REPORT RESUMES

ED 017 985 24 EA 001 176

TRAINING EFFECTS OF FEEDBACK AND MODELING PROCEDURES ON TEACHING PERFORMANCE.

BY- MCDONALD, FREDERICK J. ALLEN, DWIGHT W.
STANFORD UNIV., CALIF., SCHOOL OF EDUCATION
REPORT NUMBER BR-5-1030 PUB DATE 67
CONTRACT OEC-6-10-078
EDRS PRICE MF-$1.00 HC-$9.36 232P.

DESCRIPTORS- *TEACHER EDUCATION, *TEACHING SKILLS, *FEEDBACK, TRAINING LABORATORIES, TEACHER BEHAVIOR, VIDEO TAPE RECORDINGS, *TEACHER INTERNS, LITERATURE REVIEWS, SELF EVALUATION, SUPERVISORS, *REINFORCEMENT, MODELS, STATISTICAL ANALYSIS, LEARNING PROCESSES, QUESTIONNAIRES, EFFECTIVE TEACHING, STANFORD,

THIS REPORT DESCRIBES A SERIES OF EXPERIMENTS TO ASSESS THE USEFULNESS OF TELEVISION RECORDINGS IN IMPROVING TEACHING PERFORMANCE. OBJECTIVES OF THE STUDY ARE (1) TO COMPARE THE EFFECTS OF SELF-EVALUATION OF A TEACHING PERFORMANCE WITH FEEDBACK PROVIDED BY A SUPERVISING INSTRUCTOR, (2) TO COMPARE THE EFFECTS OF REINFORCEMENT DELAY, AND (3) TO COMPARE THE EFFECTS OF A PERCEPTUAL MODELING DEMONSTRATION OF A DESIRED BEHAVIOR WITH THOSE PRODUCED BY PROVIDING A WRITTEN DESCRIPTION OF THE BEHAVIOR AND TO COMPARE THE EFFECTS OF COMBINING REINFORCEMENT WITH EACH. EACH OBJECTIVE WAS THE SUBJECT OF A SEPARATE EXPERIMENT CONDUCTED UNDER HIGHLY CONTROLLED, LABORATORY-LIKE CONDITIONS. IT IS CONCLUDED THAT THE RESULTS OF THIS STUDY SUPPORT THE ASSUMPTION THAT THE RATE AND LEVEL OF LEARNING A GIVEN TEACHING STRATEGY VARY AS A FUNCTION OF THE MODE OF MODEL PRESENTATION. THERE IS EVIDENCE TO INDICATE THAT PERCEPTUAL MODELING PROCEDURES ARE CHARACTERIZED BY DISTINCTIVE CUING PROPERTIES WHICH TEND TO RECOMMEND THEM OVER SYMBOLIC MODELING PROCEDURES FOR USE IN TRAINING CONTEXTS ANALAGOUS TO THOSE DESCRIBED IN THE EXPERIMENT. (HW)

5-1030
24

TRAINING EFFECTS OF FEEDBACK AND MODELING PROCEDURES ON TEACHING PERFORMANCE

FREDERICK J. McDONALD
DWIGHT W. ALLEN

Stanford University
1967

EA 001 176

SCHOOL OF EDUCATION • STANFORD UNIVERSITY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TRAINING EFFECTS OF FEEDBACK AND MODELING PROCEDURES ON TEACHING PERFORMANCE

Frederick J. McDonald

Dwight W. Allen

Stanford University
1967

This research was supported by a grant from the
United States Office of Education
OE-6-10-078

INTRODUCTION

The set of experiments reported here represents an attempt to analyze the effects of reinforcement and modeling variables on the learning of a selected class of behaviors, called "technical skills of teaching." These experiments were conducted at Stanford University using the trainees in the Secondary Education Teacher Program.

We think these experiments are unique in that they are one of the few, if not the only, attempt to modify teaching behavior by using the paradigms of psychological experimentation. Each of these experiments might be seen, then, as a way of modifying teaching behavior as well as a test of the comparative effects of different kinds of procedures. Obviously, this comparison is the essence of the experimental work presented here.

The experiments also introduced certain innovations in experimental technique as applied to research on teaching behavior. Rather than use large scale experiments under regular classroom conditions, many of the experiments were conducted under highly-controlled , laboratory-like conditions. In two of the experiments, observations were made of teaching behavior in classrooms, but the actual treatments were conducted under highly-controlled conditions. We call attention to these aspects of the experiments since other investigators may want to adapt our procedures. To the degree that we have demonstrated their feasibility, others will be less reluctant to depart so radically from the usual practices in investigating teaching behavior.

We would also like to call attention to the innovative use of the videotape technology. All of this experimentation was made possible through the use of this device. Whether or not similar kinds of experiments could be conducted without it is a moot point. However, these experiments illustrate that the videotape recorder may well be the technological instrument which will substantially improve experimentation on teaching behavior.

These experiments also make a contribution to the development of psychological theory. The variables used in the experiments represent variables which are currently being investigated by a large number of researchers, though the behavior being studied is considerably more complex. To the degree that it is, these experiments represent generalizations of what has already been learned, and as will be seen when the results are read, suggest modifications of some concepts.

We would like to thank the Stanford Interns who served as the subjects in these experiments. Although the experimentation provided them with training, it nevertheless placed great demands on them. Without their assistance and cooperation the experiments would not have been possible.

We would also like to thank particularly Dr. Michael E. J. Orme, now of Harvard University, who as a graduate student served as a principal research assistant on this project. Many individuals, working as assistants, organized our data by rating videotapes and by supervising experimental sessions. Each of these persons, by his effective work made possible the successful conduct of the experiments, and deserves our gratitude even if it can be expressed in so modest a form as this acknowledgment.

We would also like to thank Dr. Robert N. Bush, who encouraged our experimentation, and who made possible its conduct in the Teacher Education Program. His continuous support of our ideas and our work was a necessary condition for its success.

Dean I. J. Quillen's support and his provision of time and resources to conduct these experiments were of inestimable value to us.

Frederick J. McDonald

Dwight W. Allen

TABLE OF CONTENTS

INTRODUCTION

Chapter I Problem and Rationale

Chapter II Experiment I: The Effects of Self-Feedback and Reinforcement on the Acquisition of a Teaching Skill 27

Chapter III Experiment II: The Effects of Feedback and Practice Conditions on the Acquisition of a Teaching Strategy 59

Chapter IV Experiment III: The Effects of Modeling and Feedback Variables on the Acquisition of a Complex Teaching Strategy 83

Chapter V Conclusions

APPENDICES

REFERENCES

CHAPTER I

PROBLEM AND RATIONALE

This report describes a series of experiments to assess the use of television recordings to improve teaching performance. The availability of portable videotape recording devices has made it possible to study teaching behavior in new ways. Contemplated uses could not be made previously either because recording equipment limited what could be done or costs were prohibitive.

Problem

One of the most difficult problems in designing instructional systems to produce teaching behavior has been to provide adequate feedback information on the teaching performance. The student or intern teacher judges for himself his own performance on a day-to-day basis. He is periodically, though not frequently, "supervised".

Usually, the character of the supervision is left to the discretion of the supervisor, even though supervisor training procedures may have been casual and sporadic. In the last analysis, it is the supervisor who observes, records, and reports to the teacher. To utilize this feedback the teacher has to see himself as somebody else has seen him, and he must relate these perceptions and judgments to his own. The difficulties in this procedure are that it invites heavy reliance on private frames of reference, the communication requires a high order of psychological skill, and it stimulates defensiveness.

All present methods of giving a teacher information about his performance have one or more of the following problems: (1) they inadequately control the defensive reactions of communicator and communicatee; (2) they require a teacher to visualize his performance from a word description of how he is behaving; (3) they do not begin from a common perception of what was done, why and how it was done, and what the effects were.

The responses of students in his class provide a teacher with another kind of feedback. This feedback, however, may be quite diverse, some students responding positively or negatively for reasons unrelated to the quality of the teaching performance. Students also mask their reactions. Test results, another form of student feedback, are explainable only in part by the effectiveness of the teaching. Although student reaction is important despite these inadequacies, the beginning teacher may be insensitive to it or may grossly misinterpret it. Even if the beginner is learning to interpret student reaction, he has difficulty relating it to the specific behavior that produced it. It seems clear that both student reaction and supervisor's comments lack the important characteristics that make a feedback process effective and efficient in producing the learning of teaching behavior.

The problem is to provide adequate feedback to control and to facilitate the acquisition of effective teaching behavior. The feedback process should have the following features: (1) the feedback process should reproduce the teaching performance as completely and as reliably as possible; (2) it should be as objective as possible; (3) it should be as immediate and as frequent as possible.

An audio-visual recording of teaching behavior has the first two characteristics. The third characteristic is acquired by the scheduling the viewing of the performance as soon after it occurs as possible and by having the recording available for frequent viewing.

Audio-visual recordings of a performance have the following additional advantages: (1) the teacher can observe his own performance; (2) when a supervisor talks to the teacher about his performance they have a common starting point. The first of these advantages presumably resolves the problem of the teacher visualizing his performance. The second advantage should tend to minimize, or at least put some "reality bounds", on the defensive reactions of the teacher and supervisor and provide for the establishment of a common frame of reference.

Another problem in the learning of teaching behavior is adequately representing the desired behavior. One way teachers learn about desired performances is by observing model teachers. These observations are usually uncontrolled; that is, the beginner may observe whatever he chooses to observe, or may not see what he has been told to observe. He may not interpret what he observes correctly. The method of observation does not provide for successive examinations of the teaching performance, nor for careful analysis of it. The same difficulties plus the effects of forgetting and selective remembering characterize the trainee's recollections of teaching performances he has observed during his own schooling. Obviously, descriptions of how to teach are fraught with all the problems of verbal description.

Audio-visual records of teaching performances may be controlled in many ways through editing. The desired display is literally created. Positive and negative instances of the behavior can be provided. The effects on student behavior can be correlated with the teaching behavior. A variety of instances demonstrated by many teachers under different circumstances can be shown. The display can be repeated and can be used in a critical analysis of the performance by instructors.

Audio-visual records, specifically videtape records, provide greater control over two processes known to be related to effective learning, demonstration of the desired behavior and feedback on a performance in which the behavior is attempted. However, basic research is needed to assess the relative importance of the two processes, their relation to the efficiency of the learning, and the influence of variables associated with each process affecting its effectiveness. The experiments described in this project are designed to assess the relative effects of various arrangements of these two processes.

Although television recordings seem to offer many advantages there is little information available on the use proposed here. We do not know how effective watching one's own performance may be, to what degree the observation needs to be structured, whether teachers' reactions to the observation will facilitate or hinder learning, and what kind of a demonstration of the desired behavior, if any, needs to be included. In short, the characteristics of an efficient demonstration and feedback process need to be determined.

Similarly we do not know how to talk to teachers about their performance when a visual record is available as the material of the discussion. Will there still be difficulty in establishing a common frame of reference? Will the teacher and supervisor even agree on what they see? Is pointing out adequate performance sufficient?

Other problems are that we do not know the contigency schedule that is most effective with this kind of feedback. Nor do we know what behavior patterns are most likely to be affected by it.

A comment on the technology of the portable television tape recorder is relevant. The portability of the recording apparatus encourages its frequent and regular use. Elaborate arrangements do not have to be made to record a class session. At Stanford we have developed a self-contained portable cart which can be installed in a classroom during the passing period between classes. Interference with classroom routine is minimized. The recording process is relatively inexpensive.

The major advantage from the viewpoint of providing feedback is that the recording is immediately available. A teacher could, for example, though this is not at present being recommended, teach a part of a class hour, step outside and watch his performance, and return to his class for another performance hopefully improved from his observations. He may also study his own or another's performance under supportive and instructive conditions.

Related Research

The paucity of research on how teachers learn to teach is best illustrated by the limited attention devoted to this topic in the Handbook of Research on Teaching (Gage, 1963). Historically, the major research method has been to correlate measures of teacher characteristics with ratings on effectiveness. Stern, summarizing the literature, notes "no substantial progress along these lines, despite the marked trend toward the use of more reliable measures of predictor variables." (Stern, 1963, p. 419). One of the major difficulties encountered in these studies has been the ambiguity of the criterion.

Another line of investigation has related specific teacher behaviors to student behavior. Flanders, studing classroom climate, has developed

THREE MAJOR CATEGORIES OF TRAINING VARIABLES

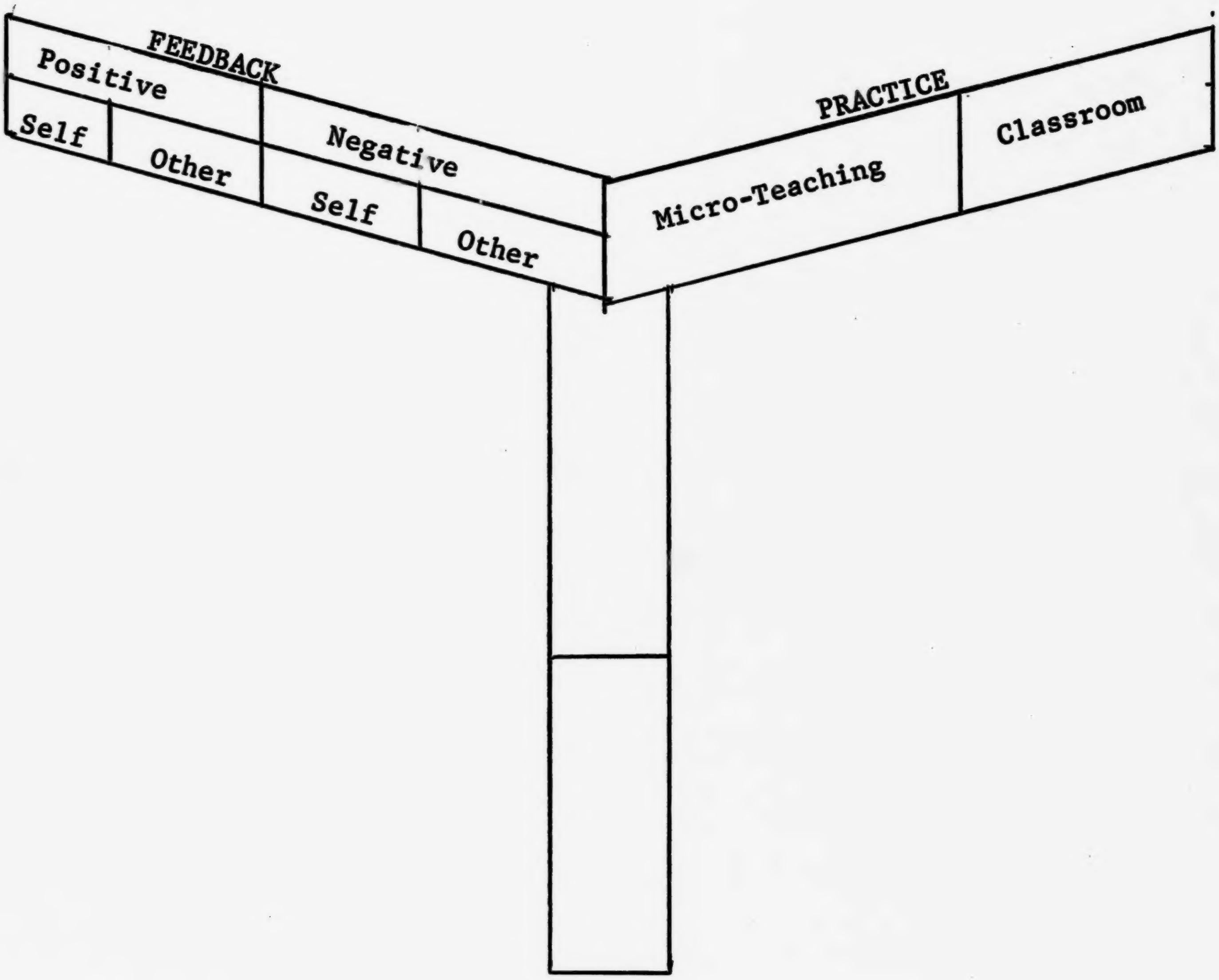
FEEDBACK
Positive
Self
Other
Negative
Self
Other
PRACTICE
Micro-Teaching
Classroom

Figure 1

Along each axis are four subcategories of variables. Thus, feedback may be either positive or negative, and may be either self- or other-administered. The proposed experimentation explores variations along each axis and the interaction among the major categories.

a recording procedure for analyzing classroom interaction (Flanders, 1960). In this method, observers classify teacher and student "talk." Flanders found that achievement was higher in classes in which teachers used indirect influence, and that students praised by the teacher received more sociometric choices. (Flanders, 1960) The purpose of the Flanders' studies was to assess interaction effects. They are part of a series of such studies (Anderson, 1945; Withall, 1949, 1956; Thelen, 1959). These studies, however, are not training studies. They, like those discussed in the first paragraph, are designed to show the relation between classes of teacher behavior and classes of student behavior. They are consistent with the notion that if certain kinds of teacher behaviors can be learned, certain effects will be produced.

Figure 1 presents a taxonomy of the variables which are relevant to the problem being studied. The major categories of variables are those associated with the feedback, practice, and demonstration or guidance systems which are the major components in training systems for learning many teaching performances.

Each of these in turn has been subdivided into the major categories being studied or proposed for study. For example, practice of a teaching performance may be conducted in either a controlled situation such as microteaching, where the teacher instructs a small group of students for a short period of time during which he attempts to enact one set of teaching responses, or he may practice the teaching skills in a classroom under the conditions prevailing there.

Feedback may be positive or negative, where positive feedback has informational characteristics, i.e., appropriate responses are indicated as appropriate; such reinforcement may also be rewarding -- "That's very good." Negative reinforcement consists in indicating inappropriate responses or omissions; such reinforcement may be mildly adversive -- "That's not a very effective way of teaching that."

Either kind of feedback may be mediated by another person, or the feedback may be mediated through a self-critique of one's teaching performance (in our experiments the use of the videotape recorder has made feasible the detailed criticism of one's own performance after the teaching event has occurred; the record is also a complete one so that the supervisor or teacher does not have to depend on his memory of the teaching performance.

The demonstration variable may also be broken down into several basic ways of portraying a desired response. The category, *symbolic demonstration*,consists in descriptions of the desired behavior, either written or spoken. In such demonstrations the subject does not view an actual portrayal of the desired behavior. The label, *perceptual demonstration* (in these studies it has also been called *perceptual modeling*)consists in portraying the desired behavior for the learner. Such portrayals may be "live" or mediated by videotape. In the experimental studies conducted in this project the modeling of desired teaching behavior has been mediated by videotapes. Each of these forms of demonstrating may be subdivided into two kinds. For example, a subject may view the best instances of his own teaching performance. This would be an instance of *self-modeling* or a *self-perceptual* demonstration. Similarly, a subject may write out for himself a detailed plan for enacting a performance which he thinks is the best possible approach to handling a problem. (We have not yet used this form, but its use seems to be feasible.) Others may give instructions (symbolic demonstration), or may portray the behavior (perceptual demonstration).

A review of the relevant literature by Bandura and Walters has shown that complex behavior may be acquired almost entirely through imitation (1963b). They state that the provision of face-to-face models serves to accelerate the learning process and in cases where errors are dangerous or costly, becomes an essential means of transmitting behavior patterns (Bandura and Walters, 1963b, p. 52). In addition, Bandura, Ross and Ross (1963c) have demonstrated that film-mediated models are as effective as real-life models in transmitting some patterns of behavior.

These experimental demonstrations of modeling effects have used young children as subjects and aggressive behavior as the dependent variable. While it has been generally speculated that similar effects occur with adults and that these effects cover a wide range of dependent variables, research literature in this respect is meager. One problem for future research is to find the modeling variables which increase the learning efficiency. It has been shown in other research that it is possible, using modeling procedures, to produce learning in which the subjects have acquired the behavior to the same level as the models (McBrearty, Marston, Kaufer, 1961).

A brief theoretical discussion of the determinants of observational learning seems appropriate. Miller and Dollard (1941) assume that such learning is contingent upon the administration of reinforcing stimuli either to the model (M) or the observer (S). Mowrer's (1960) proprioceptive feedback model also highlights the role of reinforcement in imitation learning. He distinguishes two types of such learning. In the first, S is reinforced directly. In the second, S receives vicarious reinforcement. This empathic learning, as it is termed, is assumed to occur when M exhibits responses which are reinforced and S in turn empathically experiences reinforcement as well. Hill (1960) uses this conception in generating a model of the identification process.

Bandura and Walters (1963b), however, point out that these imitation-reinforcement theories do not account for the learning of matching responses when S does not perform M responses during acquisition, or when reinforcement is not provided for either M or S. The acquisition of imitative responses under these conditions can best be accounted for in terms of contiguity theory as developed by Sheffield (1961). It is assumed that when S observes M exhibit a sequence of responses, S acquires through the contiguous association of sensory events, perceptual and symbolic responses possessing cue properties that are capable of eliciting at some time after demonstration, overt responses corresponding to those that have been modeled. On the basis of this kind of reasoning, Bandura (1963b, 1965) has suggested that the acquisition of model behaviors occurs through contiguity, while reinforcements administered to M exert their influence on the performance of imitatively learned responses. His findings support the notion that under conditions where children are exposed to aggressive models who are either reinforced, punished or left without consequences, reinforcement acts as a performance-related variable rather than a learning or acquisition one. Children in the model-punished conditions did not imitate the model nearly as often as those in the other conditions. When an attractive incentive was later offered to induce all three groups to reproduce the model's aggressive behavior, the previously observed performance differences were completely wiped out.

It cannot be assumed from the above discussion that the mere exposure of S to M constitutes a sufficient condition for imitative learning in a complex training sequence. Bandura (1964) states that factors other than contiguity undoubtedly influence imitative response acquisition.

One of the major factors associated with learning is reinforcing learning behavior (a discussion of this literature appears later in this section). Of interest is the problem of the relative effectiveness of reinforcement and modeling. The relevant research on this point is presented first.

Bandura and McDonald (1963a) have shown that under certain conditions, providing models is significantly more effective in changing behavior than are procedures in which only reinforcement is used. They designed an experiment to test the relative efficacy of social reinforcement and modeling in modifying moral judgment responses. One group of children observed adult models who expressed moral judgments counter to the group's orientation. These children were socially reinforced for adopting the models' evaluation responses. A second group observed the models but received no reinforcement. The third group did not view the models, but were reinforced for moral judgments that ran counter to their dominant evaluative tendencies. Following the treatment, subjects were tested for generalization effects. The treatments, counter to Piaget's age-specific hypothesis, produced substantial shifts in the children's moral judgment responses. As the authors predicted, modeling cues proved more effective than operant procedures. The provision of models alone was as effective as the combination of modeling and reinforcement.

As Bandura and McDonald point out, the failure of operant procedures to produce change is not surprising considering that the desired responses were very weak to begin with. In many cases the desired responses occurred so infrequently that there was little opportunity to influence them through reinforcement. Thus, even though the behaviors to be learned in a training experiment were already in the subject's repertoire, one might expect that reinforcement alone would be considerably less efficient than a procedure which was designed to highlight salient cues that preceded the desired behavior.

The problem of the studies described here is to assess the relative influence of demonstration and feedback variables on learning specified classes of teaching behaviors. The effects of one of these variables, feedback, have been extensively investigated under such labels as reinforcement, knowledge of results, effects of praise and blame, and trial and error learning. From Thorndike's formulation of the Law of Effect to Skinner's analysis of operant conditioning and despite theoretical differences, a large body of psychological literature supports the generalization that reinforcement procedures produce learning.

Knowledge of results experiments differ in the way in which the knowledge of results are obtained. In a target shooting experiment, for example, the subject obtains information directly related to correcting his performance. In experiments involving verbal learning, the reinforcement procedure not only confirms correct responses but also provides an opportunity for correction of errors and an implicit-practice trial (Michael and Maccoby, 1953, 1961).

Although these components of feedback have not always been parceled out, the results of the experimentations have been consistent; increased knowledge of results facilitates learning. However, in a complex learning task such as learning a teaching behavior, the relative effectiveness of feedback may be highly dependent on the kind of feedback provided. A teacher, for example, who watches his own performance is getting feedback, but its effectiveness may be limited compared to a confirmation procedure in which the response is confirmed and its behavioral history is discussed. Similarly, the effectiveness may vary between a supervisor confirmation and an observation of the immediate effect on student behavior.

A related issue concerns the immediacy of the feedback. Porter has pointed out that in human learning experiments a "true" comparison between immediate and delayed reinforcement has yet to be made (Porter, 1957). Because experimental control is not readily available in a classroom, the question of amount of delay of reinforcement becomes paramount. Videotape makes possible comparisons of relatively short delays to delays of hours and days. If appropriate conditions can be arranged, a teacher may step out of a class after a short teaching session and view his performance; or, he may view it at the end of the hour, or before his next teaching period, or even several days or weeks later. Each of these variations modifies the procedure so that more may be learned from the feedback. The usual inhibitory effects of delay may be diminished under these conditions. It is also possible that delay may be desirable as a means of increasing the objectivity of the teacher who may be less defensive in his analysis of performance after a lapse of time. If this is true the videotape procedures can provide a unique combination of faithful reproduction, common frame of reference, and objectivity of the feedback.

Exposing a person to a complex sequence of stimulation is no guarantee that he will attend to the entire range of cues, will select from the total stimulus complex the most relevant stimuli, or that he will even perceive accurately the cues to which his attention is directed. Sheffield and Maccoby (1960) have demonstrated that increasing the distinctiveness of relevant modeling stimuli greatly facilitates observational learning. Their work suggests that a key problem in observational learning is stimulus-discrimination learning.

According to Wulff and Kraeling (1961), one way to clarify the nature of the stimulus event in associative learning is to propose that a learner makes implicit responses to the individual features of stimulus objects (which function as stimulus events), and that those implicit responses become associated with overt responses. The stimulus features which elicit such implicit responses are termed cues.

For training research, the important implication seems to be that the treatment applied should not only require the trainee to focus on the correct end-response, but also require him to respond to salient cues that occur during the course of learning. Lumsdaine (1961) reflects this concern in suggesting that programmed learning specialists have been preoccupied with reward schedules to the detriment of the manipulation of prompting cues. Thus, undue attention has been paid to corrective feedback procedures, and the guidance of learning has been neglected.

It is obvious that the playback features of the TV unit can be used to develop procedures that emphasize cueing, since a tape can be stopped or replayed at any point in the playback. Thus both prompting and differential reinforcement techniques in discrimination training may be incorporated in training sessions. There are also implications here for testing the differential effectiveness of symbolic and perceptual modeling procedures. Before they are drawn, a word might be added concerning another aspect of the relationship between prompting and the reinforcement or confirmation elements in a training sequence.

Findings obtained by Cook and Kendler (1956) and subsequently confirmed by Cook (1958) for paired-associate learning indicate that when a prompting procedure is pitted against a no-prompting confirmation procedure, the former is superior. Angell and Lumsdaine (1960) have obtained evidence that a mixture of prompted and unprompted trials is more effective than exclusive use of prompted trials. More recently, these authors have demonstrated that when adequate prompting is provided, variations in the kind of confirmation or feedback given are much less important than when less prompting is provided (Angell and Lumsdaine, 1961).

What are the implications of this research for studies of the effects of symbolic and perceptual modeling? In general, perceptual modeling should be superior to symbolic modeling in a videotape training session. It is reasonable to assume that in a perceptual modeling condition, the experimenter would have more opportunities to provide discrimination training on relevant cues since the model could be trained to emit the criterion behavior frequently. Further, in the perceptual modeling condition there is relatively greater emphasis on prompting techniques.

Having viewed the prepared tape beforehand, E could say to the subject, "Look, the pupil is about to respond (say, by asking a question).

The model is going to give both non-verbal and verbal reinforcement for this." In the symbolic modeling procedure, greater emphasis is placed on confirmation techniques since E and the trainee view the latter's playback, not a prepared model tape.

Finally, Sheffield's (1961) discussion of perceptual blue-printing would lead one to predict that a relatively greater emphasis on viewing a model over viewing the self would produce the greatest increase in response strength for a given class of behaviors. This perceptual blueprint notion is like Tolman's (1959) conception of a cognitive map in that both serve to organize discrete stimulus-response elements in a given task. The advantage of such an organized sequence is that it provides a distinctive stimulus context within which appropriate response items can be specifically cued (Sheffield, 1961, p. 29).

This means that in a training situation where the responses required are already part of the subject's repertoire, the problem becomes one of increasing the saliency of the perceptual aspects of the treatment The perceptual modeling condition appears to be superior to the symbolic modeling condition in this regard because the model teacher can be trained until clear-cut and unambiguous displays of the criterion behavior are filmed. This display includes the verbal and non-verbal cues that signalled the occurrence of the criterion behavior. In this context, perceptual adequacy is of prime importance.

Other things being equal, perceptual modeling should be a more efficient training procedure than symbolic modeling for the following reasons: perceptual modeling procedures permit one to display a large number of the desired responses; they increase the distinctiveness of relevant stimuli; and in training sessions E uses them to highlight cues and responses by prompting techniques. Lastly, it is suggested that the perceptual adequacy of such modeling facilitates the development of perceptual blue-prints which serve to unify discrete elements of the desired skill.

The major issue is whether feedback needs to be as extensive if relevant response demonstration is available. The theoretical issues and related research have recently been analyzed by Bandura and Walters (1963b). In a series of experiments, Bandura has shown that the availability of models portraying a behavior class significantly influences behavior change. Bandura and McDonald have further shown that experimental conditions where models were present significantly changed behavior over the condition where only reinforcement was used in behavior shaping (Bandura and McDonald,1963).

One reason advanced for the superiority of the modeling conditions is that it eliminates the relatively lengthy period of trial and error and shaping of successive approximations. Another reason is that the model's behavior heightens awareness of the relevant cues to which responses are to be attached, again eliminating the trial and error period. There is also little ambiguity about either the character of the response or its effectiveness.

In demonstrating teaching behavior, it would be expected that the effects of modeling would be heightened. The cues are more diverse and more subtle. The desired response is not easily evoked partly because of its complexity, partly because the learner has never initiated the response under a similar set of conditions. The model is known to be portraying the desired behavior, a condition not explicitly present in other experiments.

The experimental literature suggests that both modeling, a form of demonstration, and feedback will be effective. But as noted earlier, their relative effectiveness needs to be assessed. The human learner because of his greater information-processing capacities may require less feedback when the desired response is modeled. The feedback may be more effective when delayed because greater use can be made of this information-processing capacity. However, the responses to be learned may be of sufficient difficulty to require some combination of demonstration and feedback, to need both a portrayal of the desired response and a period of corrected practice.

The studies described here are experiments to assess these relative effects and to study some of the variables associated with both processes. Both their theoretical and practical significance is considerable. A feedback procedure can be expensive in time and materials. If the information-processing capacities of the human learner can be used to shorten this procedure, a finding in this respect would be highly useful in developing an instructional system for teacher training. Determining the most effective forms of feedback would be similarly useful. Such results would also contribute to the analysis of the specific contribution of feedback procedures in human learning.

The Hunter College Study of the Improvement of Student Teaching utilized television recordings to improve the quality of instruction given by student teachers. (Schueler, et al., 1962) The television record did not significantly improve student teaching when used as a part of the supervisory process. The authors of the study discuss several reasons why results probably were not attained, among them the range of variation in the teaching situations. Although this project has been successful in developing a meaningful and reliable measuring instrument, the procedure sampled a relatively large range of teaching behaviors. The extensiveness of the changes required may not have been achievable in the period of student teaching, at least not in sufficient amount to yield significant differences. Similarly, for a large sample of behaviors, carefully described, the visual record may not add enough to produce significant changes. Although verbal descriptions are not usually satisfactory, the instrument used in this study describes behavior specifically, and the behavior frequencies are tabulated for the inspection of the student teacher. The problem seems to be to determine what the visual record adds to the learning conditions.

Objectives

1. To compare the effects of self-evaluation of a teaching performance with feedback provided by a supervising instructor. (Experiment I, see below) Does a learner need the cueing properties provided by an observer when his own performance is displayed to him?

2. To compare the effects of delay of reinforcement and the kind of reinforcement provided (Experiment II, see below) The effects of delay may be reduced when the feedback is more comprehensive.

3. To compare the effects of a perceptual modeling demonstration of a desired behavior with those produced by providing a written description of the behavior (symbolic modeling), and to compare the effects of combining reinforcement with each. (Experiment III, see below) The

desired behavior may be learned more easily from written descriptions; however, most theorists believe that observing the model is necessary to learn the cues to which the desired behavior is attached.

The experiments described attempt to answer three questions which reflect these objectives: (1) What is the relation between a modeling demonstration and a feedback procedure to the acquisition of a teaching skill? (2) What is the effect of varying the source of feedback, self or other, provided? (3) What is the effect of varying the amount and kind of feedback?

Procedure

A series of experimental studies are described here, each in detail below. These studies have several characteristics in common. Each utilizes a before-after design. The base rate of the behavior to be learned is measured in pre-experimental videotape recordings. This measurement is necessary since the desired behavior is likely to be already occurring in some form and at some frequency level.

Each experiment limits the behavior class to be learned. A specific behavior, for example, reinforcing student participatory responses, is the behavior to be acquired under one set of experimental conditions. This limitation effectively eliminates the problem of attempting to change a wide range of teaching behaviors and of measuring these changes under the rubric of general teaching effectiveness.

Experiment 1: A Comparison of Self-Evaluation and Social Reinforcement on the Acquisition of a Teaching Behavior.

a) Teaching interns were randomly assigned to one of four experimental conditions. The independent variable was Self-Other evaluation of the teaching performance. The dependent variable was the frequency of the desired behavior, approval of participating response by students.

In the first condition, the teaching interns viewed their own performance, counting the number of approving and disapproving responses. In the second condition, a supervising instructor viewed the intern's recording with him, and reinforced each desired response, by general approval, as it occurred. In the third condition, the supervisor provided general approval but also pointed out the effects of the teacher's approval on student participation and suggested ways of improving it. This arrangement was feasible because of the playback feature of the videotape mechanism. In the fourth condition, the control group, subjects viewed their own performances but were given several categories of teaching behavior to analyze.

The rationale for these treatments was that the supervisor-evaluation conditions cued the subject to the character of the desired response, and in the third condition pointed out the consequences of the teacher behavior. These conditions provided specific and controlled feedback. A competing hypothesis was that if the subject was directed to observe and count his own behavior he would provide his own feedback. The experiment provided a test of these two competing hypotheses.

Three evaluation sessions were run for each subject, and the time interval between viewing and the next recording was held constant. Subjects in the control group were also recorded three times with the same time intervals intervening.

b) The sample of subjects were first year graduate students enrolled in a teaching intern program at Stanford University. All were beginning teachers.

c) Data gathered: A pre-experimental, post-experimental, and two rexperimental recordings were made of the teaching behavior. All recordings were made in the interns' classrooms. Each of these records was analyzed by pairs of independent observers who did not know the treatment given each subject. These observers counted the number of the desired responses. This number was then treated as the score for that session.

d) Analysis: A mixed model, analysis of variance design was used to analyze the data, with rows standing for individuals, and columns for treatments, and each cell containing the replications. The .05 level of significance was chosen for rejecting the null hypothesis.

Experiment II: Effects of Feedback and Practice Conditions on the Acquisition of a Teaching Strategy

The treatment of supervision sessions was alike for all subjects in that they viewed videotaped playbacks of their earlier teaching performance with an experimenter who provided discrimination training. However, instead of varying the amoung of feedback within each session as in the first experiment, we held within-session feedback constant, and manipulated the amount of practice and delayed feedback over four experimental groups. In addition, the post-test was videotaped approximately seven weeks after training so as to permit inferences about retention as well as acquisition curves.

The teaching behavior to be learned in this experiment was a redirecting and refocusing technique for handling students' questions. The purpose of the technique was to encourage the student to answer a question for himself or to correct his response, in contrast to giving an answer or correcting one.

Three experimental sessions were run for each subject with time intervals between viewing and recording sessions controlled. Subjects were randomly assigned to conditions:

b) Sample of subjects: In all of these experiments the same sample of subjects were used.

c) Data gathered: Again, pre-experimental , post-experimental, and two experimental sessions were recorded. Independent observers counted the number and kind of question-handling techniques. The number of each kind was used as a score for each individual.

d) Analysis of data: As in Experiment I.

Experiment III: The Effects of Modeling and Feedback Variables on the Acquisition of a Complex Teaching Strategy.

One of the objectives of the study was to compare two modes of model presentation in training sessions for intern teachers. The experimental design tested the hypothesis that the rate and level of learning a given teaching skill varies as a function of the mode of model presentation. Two types of modeling were considered: (1) Symbolic Modeling: This is defined as a process whereby one transmits desired behaviors to the learner by means of written or verbal instructions. The subject does not view an actual portrayal of the desired behavior.

(2) Perceptual Modeling: This is defined as a process whereby one

transmits desired behaviors to the learner by means of a filmed model who portrays the desired behavior.

Three experimental sessions were run for each subject. Subjects were assigned randomly to treatments.

a) A new sample of subjects, the second class were used in this experiment. These subjects had the same general characteristics as the subjects described in the first two experiments.

b) Data gathered: The same general procedures described in the preceding experiments were used in this experiment. The behavior class of interest is the probing behavior described in Experiment II

c) Analysis of Data; As in Experiments I and II

General Considerations

1) In each of these experiments a single teaching behavior was the object of study. No brief is made that these were the most important skills to be learned, nor that their acquisition would guarantee teaching success. They were chosen because they represented generalized skills that a teacher probably needs to have in his repertoire. Experience with beginning teachers has indicated that they have difficulty in performing these behaviors. They typically give more disapproving responses than approving responses to student participatory behavior, are more likely to answer a question than prompt a student to answer it, or direct the question to another student. They tend to confine their questioning and prompting to a limited number of students.

What has been learned from these experiments about the acquisition of these skills may be generalized to similar skills. The variables that affect the learning of the simple form of the skill presumably would influence the learning of the more complex variation of the skill.

2) A portable videotape recorder produced by Machtronics, Inc., Palo Alto, California was used. This instrument has been used for two years in pilot studies and for clinical training. The equipment can be set up in a classroom between periods, and operated quietly. Its small size renders it unobtrusive.

3) Both teachers and classes were run through a familiarization period before the actual filming of experimental recordings. Experience has indicated that students largely ignore the filming when the equipment has been present for several days. The interns became accustomed to using the recordings during the summers preceding the experiments. During this period they became accustomed to seeing themselves on videotape.

4) Length of viewing samples was standardized for all experiments. Conditions for viewing were also standardized. Experience with the recorder has familiarized the camera crew with desirable recording angles, which were also standardized.

5) Reliability of scoring was computed from the ratings of the independent observers. A training period was provided with films already available. Experience with these ratings yielded high observer agreement.

6) Specific directions were given to experimental subjects on the general characteristics of the lesson to be recorded. These directions precluded recording study sessions or hour lectures with no student participation.

In the following chapters, each of the experiments is described in detail. The design and results of each experiment are presented separately. Tables presenting the data accompany the description of the results. Materials used in the experiment, such as directions given trainees, may be found in the back of this report together with additional statistical information.

CHAPTER II

EXPERIMENT I

THE EFFECTS OF SELF-FEEDBACK AND REINFORCEMENT ON THE ACQUISITION OF A TEACHING SKILL

Materials for this study may be found in Appendices A-E.

The experiment described here applies well-known principles of reinforcement theory to a training problem. The training paradigm involves applying a reinforcer to an emitted response. The prediction is that the rate of responding will increase.

Two aspects of the training procedures do, however, depart from those typically used in reinforcement studies. First, the reinforcer is not given while the learner is actually emitting the response. In this study, the subjects were videotaped while emitting complex responses (teaching) some of which were to be reinforced. After the actual behavior sample was collected the subject viewed his performance in the presence of an experimenter. When the desired responses appeared on the videotape, the experimenter reinforced their occurrence If results similar to those obtained in other studies occur in this situation, reinforcement concepts are widely applicable. Also training procedures or complex skills can be developed which use these principles and concepts in ways directly analogous to the procedures used in laboratory studies where the utility of these concepts has been amply demonstrated.

The second characteristic of the training procedure which departs from the usual laboratory methods is that the behaviors to be learned occur in the context of many other behaviors and are relatively more complex than operants conditioned in laboratory studies. They are more analogous to the kinds of behaviors that have been verbally conditioned in psychotherapy sessions. Here, as in these other complex verbal interaction, it is literally impossible to reinforce every instance of the operant being conditioned. The immediacy of the reinforcement, for similar reasons, is also somewhat variable. Again, if comparable results are obtained, the generalizability of reinforcement concepts is supported.

Since the subjects in this experiment are humans it was also possible to test the efficacy of a cue-discrimination procedure This procedure consisted of pointing out to the subjects those cues to which the reinforced operant was attached. In this way the cue-response chain is clearly indicated which should be facilitating.

Since human subjects are being used, it is also possible that they can reinforce themselves or, more generally, provide their own feedback. Videotape recordings of a trainee's behavior sequence can be given to him so that he can view himself as a behaving organism. With instructions he can note the presence or absence of the desired response. However, it is dubious if such a procedure is likely to be highly facilitating since the subject may not attend well, may be easily distracted, may be highly subjective in his viewing. The effectiveness of this training procedure remains to be tested, however.

This experiment, then, tests the relative effectiveness of three training procedures, each representing an application of reinforcement principles. The three procedures represent points on a continuum from self-administered feedback or reinforcement to experimenter-administered reinforcement with cue-discrimination training

The prediction is that the order of training effectiveness will be in the same direction, with the self-administered feedback the least effective and the experimenter-administered feedback with cue-discrimination training the most effective.

METHOD

General Procedure: Intern teachers were videotaped on four separate occasions during the first 20 minutes of regular classroom lessons. In the intervals between each of these taping sessions they received differential feedback as part of their regular supervision. The treatment or supervision sessions were alike for all subjects in that they viewed videotape playbacks of their earlier teaching performance. The mode and amount of feedback given each intern was varied by manipulating the reinforcement and discrimination training provided by an experimenter.

Before the pretest videotapes were recorded, all subjects were told when they would be taped and were asked to present a discussion-type lesson in which teacher-pupil interaction could be observed. This was not a new or unusual experience for them as they had been frequently exposed to the videotaping-playback-supervision process during the previous three months of the Stanford Intern Program.

Pupils in each of the classrooms were informed by the interns before hand that the portable TV equipment would be present in the room, and that the cameras would be focused on the teacher, not the class.

Treatments: Mode of feedback, type of reinforcement and amount of discrimination training were varied for four experimental groups of interns.

Controls: Group (C): At the beginning of the first playback session these subjects were given written instructions which suggested that as they viewed subsequent playbacks of themselves, they try to determine their effectiveness in relation to: the aims of the lesson; use of examples; effectiveness of teacher-questions; amount of pupil participation; pacing of the lesson; and teacher-pupil rapport. Following this, they viewed the first and all subsequent playbacks alone. E started the machine and left the room. He returned as the tape finished, stopped the machine and told the intern when to expect the next taping and the date for the following playback session. As for all subjects, Group C (controls) viewed playbacks of the preceding lesson within three days of its taping. Lessons were videotaped within two days of each playback session.

Self-Feedback: Group (S-F): These subjects followed the same basic schedule as the controls, except that they received a different set of written instructions. The instructions discussed the educational relevance of increasing pupil participation in certain types of lessons; defined pupil participatory responses (PPR) as a clearly observable non-verbal or verbal response that was considered desirable; and provided brief examples of such behavior. It was also suggested that the intern immediately reward a PPR when it occurred as this would tend to increase pupil participation. Examples of teacher responses - both verbal and non-verbal- were then provided. Finally, a simple rating chart was attached so that the intern could classify his responses to PPR-s as "teacher rewards"; "teacher ignores"; "teacher punishes"; or "can't classify." As with the control group, self-feedback subjects viewed each playback alone.

Reinforcement-Only Subjects: Group (R): These interns received the same written instructions as those in the self-feedback condtion. However, E viewed each of the three videotaped playbacks with them, and verbally reinforced all observable instances of these Group (R) interns reinforcing PPR's. Whenever the intern was observed to reinforce a PPR, E responded by saying "Good!"; "That's it!"; etc. Beyond this, he did not comment upon the intern's teaching performance.

Reinforcement Plus Discrimination Training: Group (R + D): These interns were first given written instructions which were identical to those administered to Group (S-F) and Group (R) subjects. In viewing subsequent playbacks with them, E provided differential reinforcement as in the Reinforcement-Only treatment. In addition, he provided discrimination training. This consisted of pointing out salient cues to which reinforcement should be attached, suggestions related to the immediacy, affect-loading and types of reinforcement the teacher could use, and finally, the effects of such behavior upon pupil participation. In general terms, then, it might be said that Group (R + D) subjects received "maximum supervision" and Group (S-F) subjects received "self-supervision".

Experimental playback sessions for all groups were thirty minutes in length. When E viewed playback with Group (R) and (R + D) subjects, he reduced the sound momentarily when providing discrimination training, and spoke over the tape when providing reinforcement. Interns in all groups were informed that they could have the tape stopped or reversed and played over again at any point during the playback. Playbacks were stopped occasionally by E when he was working with the supervised groups. However, since the sessions were limited to thirty minutes, only one or two brief stoppages in sessions two and three were possible.

Subjects:

All Stanford intern teachers preparing to teach English, social studies and mathematics were included in the study. Approximately equal numbers of interns from each of the subject-matter areas were assigned to each group. In addition, groups of interns teaching in the same school were distributed throughout the four groups. In this way, systematic bias due to subject-matter major or pupil characteristics based on socio-economic status was avoided. The major characteristics of the sample are summarized in Table 1. (See Table 1, Experiment I).

The Dependent Variable:

The dependent variable was defined as the relative frequency with which the teacher positively reinforces pupils' participatory responses during teacher-pupil interaction in the classroom. PPR's were defined in training sessions with the interns as any desirable or relevant pupil comment, answer or question. For purposes of measurement however, desirability and relevancy were not considered.

The basic strategy in defining the dependent variable involved classifying teacher responses into one of four major response categories. These include positive reinforcement, negative reinforcement, interaction and information-giving responses. Pupil responses were also classified and considered in relation to teacher responses. Each of the above response categories are defined in the following discussion. A summary of the classification system appears in Table 2. (See Table 2, Experiment I).

TABLE 1 (EXPERIMENT I)

MEAN AGE AND FREQUENCY DISTRIBUTIONS FOR SEX, GRADE-LEVEL TAUGHT, AND SUBJECT-MAJOR FOR EACH EXPERIMENTAL GROUP

N = 71

Experimental Group	Mean Age	Sex		Grade Level		Subject-Major		
		Male	Female	9-10	11-12	English	Soc.St.	Math
Group 1 Controls (N = 18)	24.9	4	14	13	5	9	4	5
Group 2 Self-Feedback (N = 18)	24.6	5	13	13	5	7	7	4
Group 3 Reinforcement Only (N = 18)	24.4	5	13	11	7	5	9	4
Group 4 Reinforcement + Discrimination Training (N = 17)	24.0	3	14	12	5	8	8	1

TABLE 2

(Experiment I)

THE DEPENDENT VARIABLE DEFINED IN TERMS OF SELECTED RESPONSE CATEGORIES

Teacher Responses		Teacher - Pupil Interaction	
Positive Reinforcement	Negative Reinforcement	Pupil Initiated	Teacher Initiated
+ Verbal Reinforcement (+VR) + Non-Verbal Reinforcement (+NVR) + Qualified Reinforcement (+QR) Post Hoc Reinforcement (PHR)	- Verbal Reinforcement (-VR) - Non-Verbal Reinforcement (-NVR) - Qualified Reinforcement (-QR)	Volunteer Comments (V) Volunteer Questions (V?) No Response (PNR)	Direct Question to one individual (DSI) Direct Question to Group, then Individual Specified (DSG-1)
Teacher No Response (TNR)		Teacher question directed to the group followed by volunteered pupil response (DSG-V)	

Teacher Positive Reinforcement: A teacher response was defined as positively reinforcing if it met one of two conditions: First, the response had to immediately follow a PPR. Secondly, it had to be classifiable under one of the following response categories. (1) Teacher Positive Verbal Reinforcement (+VR): Immediately following a PPR, the teacher uses words and phrases such as "Good," "Fine!" (2) Teacher Positive Non-Verbal Reinforcement (+NVR). The teacher in responding to a PPR, nods, smiles, leans or moves toward the pupil, or writes the pupil's response on the blackboard. (3) Teacher Positvely Qualified (+QR) and (4) Post Hoc Reinforcement (PHR). The teacher emphasizes positive aspects of pupil responses by reorienting class attention to earlier contributions by a given pupil (PHR), or by differentially reinforcing the acceptable components of a partially adequate response (+QR).

Teacher Negative Reinforcement: A teacher response was defined as negatively reinforcing if it immediately followed a PPR, and was classifiable as the obverse of one of four types of reinforcement outlined above (-VR; -NVR; -QR).

Teacher-Pupil Responses Independent of Reinforcement Classifications: Certain responses that occur frequently in interaction and yet are not classifiable as some form of reinforcement were included in the definition of the dependent variable. These included information-giving by the teacher, teacher-initiated interaction (i.e., questions directed to a given pupil or to the class in general), pupil initiated interaction (i.e., volunteered comments or questions), teacher-no-response and pupil-no-response.

In general, the dependent variable included evaluative and informational signals which the teacher may use in the classroom. There was a tendancy to emphasize socially rewarding operants since it could be expected that such behaviors would tend to increase pupil participation.

Both in the training and measurement phases of the study, the PPR was presented as an S^D which served to cue the teacher or rater that a desirable (or classifiable) teacher response was about to occur.

Measurement Procedures: Four videotapes for each intern in each group were analyzed by raters trained for this purpose. Throughout the rating phase of the experiment, they worked on the tapes in a random order so that they neither knew the treatment condition nor the number of the teaching trial of the tape being rated. Operators ran the television equipment and selected tapes using a list of random numbers.

In addition to recording the frequencies of each of the behaviors defined as components of the dependent variable, the raters recorded other relevant behaviors and lesson characteristics as well. Frequencies were recorded for the total number of pupils who responded, the number of responses they emitted, and the sex of each responder. The raters also recorded the length of each videotape to the nearest tenth of a minute, and determined how much time was spent in discussion, group work or individual study.

The Unit of Measurement: In analyzing pupil-teacher interaction, one may record discrete responses, or measure in terms of some unit such as the uninterrupted utterance. Raters were trained to define an interruption as a comment or question. "Partial" responses that teachers commonly emit during pupil speech ("um-hum"; "yes," etc.) were not defined as interruptions. All forms of verbal and non-verbal reinforcement were scored in terms of discrete responses. For example, if the teacher said, "Good!"; "Good!"; "That's fine'", the rater coded all three operants.

Training of Raters: Eight raters were initially given intensive training on intern videotapes. Once they had achieved at least 90% interrater agreement on all of the major response categories, and better than 95% agreement on teacher reinforcement responses, the analysis of experimental tapes was begun. Reliability was maintained throughout the analysis by scheduling frequent joint rating sessions where raters checked the percentage of agreement and referred to definitions of relevant responses so that systematic rating biases would not develop. Neither ratings taken during training, nor those produced in the joint sessions were used in the statistical analysis of results.

The ratings upon which the reliability coefficients reported in Table 3 are based were acquired in the following way. As each block of 30 or 40 tapes were completed (a total of 269 were actually rated), each of the six raters who did the bulk of the rating then rated a given tape. This was done without the rater's knowledge. Eight tapes, two from each trial, and two from each group were rated by all six raters in this way. As can be seen, interrater agreement is high. (See Table 3).

A certain amount of data was lost between initial videotaping and the final statistical analysis. Some tapes were technically poor, and could not be rated. Some tapes were inevitably less than the required 20 minutes, and were also omitted. However, if a given tape was over

TABLE 3 (EXPERIMENT I)

INTERRATER RELIABILITY FOR SIX RATERS
ON THE MAJOR RESPONSE CATEGORIES
OF THE DEPENDENT VARIABLE

Response Category	Kendall Coefficient of Concordance: W
Positive Verbal Reinforcement	1.00
Positive Non-Verbal Reinforcement	0.99
Negative Verbal Reinforcement	1.00
Negative Non-Verbal Reinforcement	0.87
Total Pupil Reinforcement	0.97

15 but less than 20 minutes, the obtained ratings were prorated to a 20 minute base. Of the original set of 284 videotaped lessons, 25 were omitted at the outset, and 51 tapes were prorated before statistical analyses were performed. The T statistic was used to determine whether or not certain cells in the matrix were biased by the inclusion of a disproportionate number of adjusted tapes. T was non-significant. Omitted and prorated tapes are shown by group and trial in Table 4. As can be seen, they are scattered throughout groups and trials. (See Table 4, Experiment I.)

TABLE 4 (EXPERIMENT I)

NUMBER OF VIDEOTAPES OMITTED
AND PRORATED BEFORE STATISTICAL ANALYSIS

	Trial 1		Trial 2		Trial 3		Trial 4		TOTALS	
Group	P	O	P	O	P	O	P	O	Prorate	Omit
1 (C)	2	2	2	2	1	0	2	1	7	5
2 (S-F)	3	2	4	1	3	0	1	2	11	5
3 (R)	6	1	1	1	5	1	6	4	18	7
4 (R + D)	4	4	2	2	4	1	5	1	15	8
TOTALS	15	9	9	6	13	2	14	8	51	25

RESULTS

Three types of analysis were performed upon the data. Analyses of covariance were employed to test for the significance of differences between each of the groups (treatment differences). In addition, T tests were used to determine the significance of differences within a given group from one trial to the next (training differences). Finally, a multiple regression analysis was performed on all of the major response categories to determine significant relationships among these teacher-pupil behaviors.

Treatment Differences: Positive teacher reinforcement constitutes the major response category of the dependent variable. Using trial one scores as covariants, the groups were found to be significantly different from each other. The data in Table 5 summarize these results, and show that the differences were significant for both positive verbal ($p=.001$, .005 and .025 for trials 2, 3, and 4 and non-verbal reinforcement ($p=.025$ and .005 for trials 2 and 4). (See Table 5, Experiment I)

When the two types of positive reinforcement are taken together, it can be seen that the R + D group outperformed all other groups. These relationships are illustrated in Figure I. (See Figure I, Experiment I) Positive verbal and non-verbal teacher reinforcements are presented in terms of adjusted mean frequencies for all trials and groups. Figure 2 presents the change for positive verbal and non-verbal separately.

Negative verbal and non-verbal teacher reinforcement occurred infrequently throughout the groups. Table 6 shows that all three of the experimental groups consistently emitted fewer negatively reinforcing responses.

TABLE 5 (EXPERIMENT I)

SUMMARY OF THE ANALYSES OF COVARIANCE FOR THE EXPERIMENTAL GROUPS, WITH TRIAL ONE SCORES AS COVARIANTS AND TEACHER POSITIVE VERBAL AND NON-VERBAL REINFORCEMENTS AS THE DEPENDENT VARIABLES

		Group 1 (C)	Group 2 (S-F)	Group 3 (R)	Group 4 (R+D)	df	F
		Adjusted Means and Standard Errors					
Trial 2	Verbal	30.51 4.4678	30.18 4.3160	45.18 4.2841	60.93 4.2532	3/60	11.884***
	Non-Verbal	11.50 3.1517	9.39 3.0566	13.54 2.9530	22.13 2.9556	3/60	3.482*
Trial 3	Verbal	18.61 4.231	28.83 4.342	33.46 4.272	50.66 4.243	3/62	12.140***
	Non-Verbal	13.97 2.8720	7.12 3.9639	12.40 2.8627	16.62 2.86 8	3/62	1.858
Trial 4	Verbal	27.00 5.319	23.51 5.283	26.06 6.323	47.21 5.146	3/52	4.318*
	Non-Verbal	6.67 2.3041	7.90 2.2954	9.51 2.6702	19.28 2.2145	3/52	6.561***

* .025 level of significance

** .005 level of significance

*** .001 level of significance

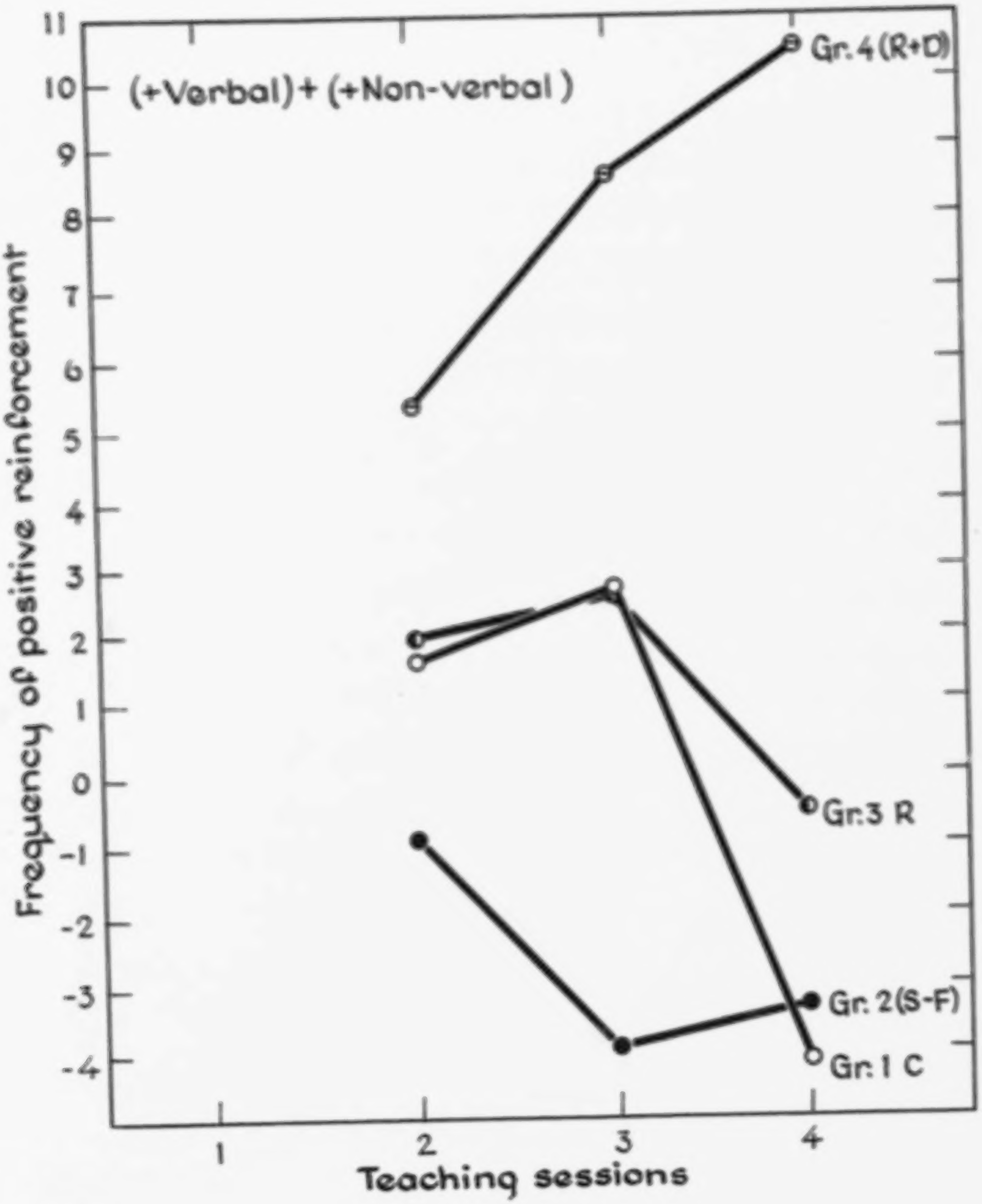

Figure 1, Experiment I

Changes in Frequency of Positive Verbal and of Non-Verbal Combined for Each Experimental Group

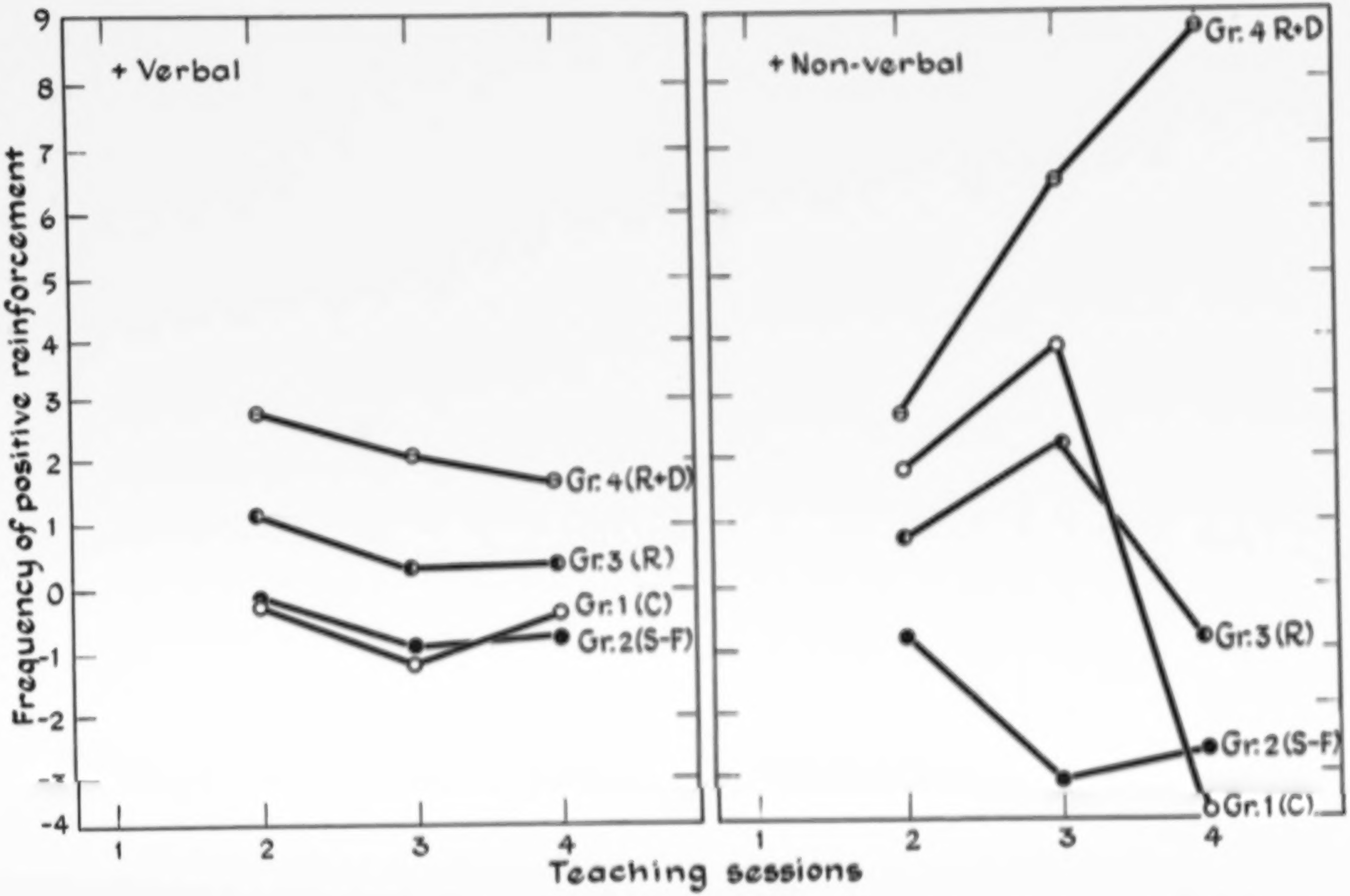

Figure 2, Experiment I

Changes in Frequency of Positive Verbal and Positive Non-Verbal Reinforcements for Each Experimental Group.

over trials while control group responses increased. These differences appear to be fairly stable, but do not reach an acceptable level of significance. (See Table 6, Experiment I) Figure 3 illustrates group trends in negative reinforcement. The data here are also based on adjusted treatment means. (See Figure 3, Experiment I)

Training Differences: Table 7 summarizes the significance levels obtained when within-group treatment means were compared using the T statistic (Winer, 1962, pp. 55-60). As can be seen, Group (R + D) interns significantly increased their rate of positive reinforcement by trial 2 ($p=.01$). (See Table 7, Experiment I) They increased from a base rate of reinforcing approximately 60% of all PPR's to a rate of 76% by trial 2. In trial 4 the rate dropped to 67% and this combined with considerable variation within the group, produced before and after treatment differences that were non-significant ($p=.10$). However, when trial 1 versus trial 4 differences were tested on the assumption that the population variances were unequal, significance well beyond the .05 level was obtained (required $.05 = 2.13$; obtained $T = 4.19$, Ferguson, 1959,pp. 143-145). Group (R + D) subjects also increased their rate of positive non-verbal reinforcement from trial 1 to 2 and following two treatment sessions, ended ($p=.10$) to use less negative verbal reinforcement.

Increases in mean positive verbal reinforcement and a concomitant drop in negative reinforcement can be most clearly seen in Group (R) subjects. The higher significance levels result from considerably less within-group variation.

The control group showed no significant within-group shifts in

TABLE 6 (EXPERIMENT I)

SUMMARY OF THE ANALYSES OF COVARIANCE FOR THE EXPERIMENTAL GROUPS, WITH TRIAL ONE SCORES AS COVARIANTS AND TEACHER NEGATIVE VERBAL AND NON-VERBAL REINFORCEMENTS AS THE DEPENDENT VARIABLE

		Group 1 (C)	Group 2 (S-F)	Group 3 (R)	Group 4 (R+D)	df	F
		Adjusted Means and Standard Errors					
Trial 2	Verbal	5.15 1.4149	1.88 1.3759	5.53 1.3392	5.52 1.3287	3/60	1.663[a]
	Non-Verbal	1.7069 4.695	1.0716 4.863	.5838 4.701	1.6419 4.703	3/62	1.264
Trial 3	Verbal	4.74 1.5257	5.37 1.4327	3.87 1.3940	6.06 1.3838	3/59	0.446
	Non-Verbal	1.51 0.5384	1.07 0.5054	0.58 0.4887	1.81 0.4899	3/59	1.184
Trial 4	Verbal	5.77 1.3748	3.39 1.3790	2.61 1.6164	3.29 1.3304	3/52	.941
	Non-Verbal	1.57 0.5060	0.94 0.5127	0.84 0.5491	0.75 0.4623	3/52	0.535

[a] .250 level of significance

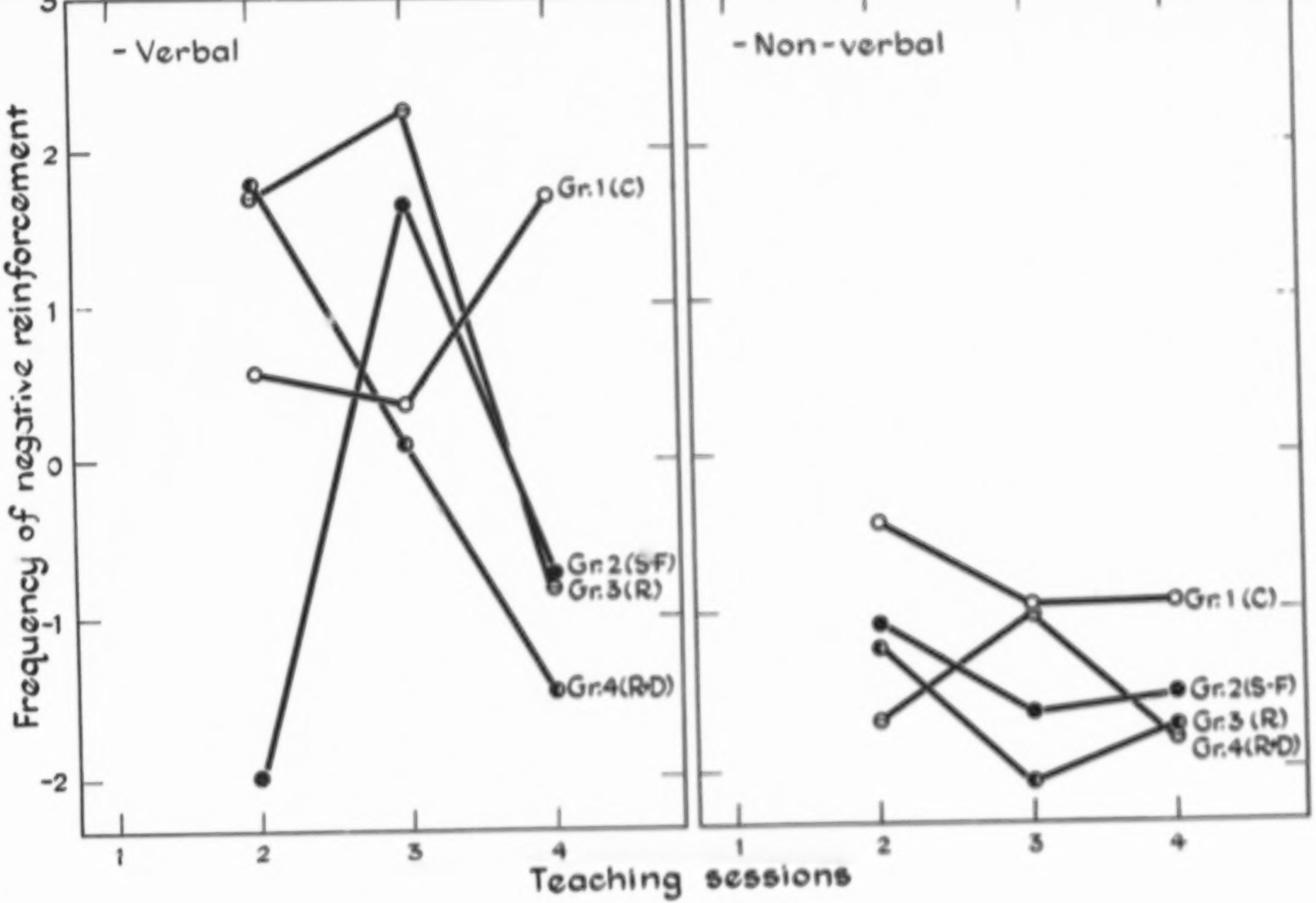

Figure 3, Experiment I

Changes in Frequency of Negative Verbal
and Non-Verbal Reinforcements for Each
Experimental Group.

TABLE 7 (EXPERIMENT I)

SIGNIFICANCE OF DIFFERENCES FOR THE FOUR EXPERIMENTAL GROUPS FROM TRIAL TO TRIAL, ON MAJOR RESPONSE CATEGORIES OF THE DEPENDENT VARIABLE*

Response Category	Group 1 (Control)	Group 2 (S-F)	Group 3 (R)	Group 4 (R+D)
	Direction of Difference from Trial to Trial and Level of Significance			
+VR	NS	NS	$T_1<T_2$ (.01) $T_4<T_2$ (.05)	$T_1<T_2$ (.01) $T_1<T_4$ (.10)
+NVR	NS	NS	NS	$T_1<T_2$ (.10)
-VR	NS	NS	NS	NS
-NVR	NS	$T_1>T_3$ (.10) $T_1>T_4$ (.10)	$T_4<T_1$ (.05) $T_3<T_1$ (.05)	$T_4<T_1$ (.10)

*Significance levels were tested by the T statistic for comparisons among treatment means (Winer, 1962, pp. 65-70).

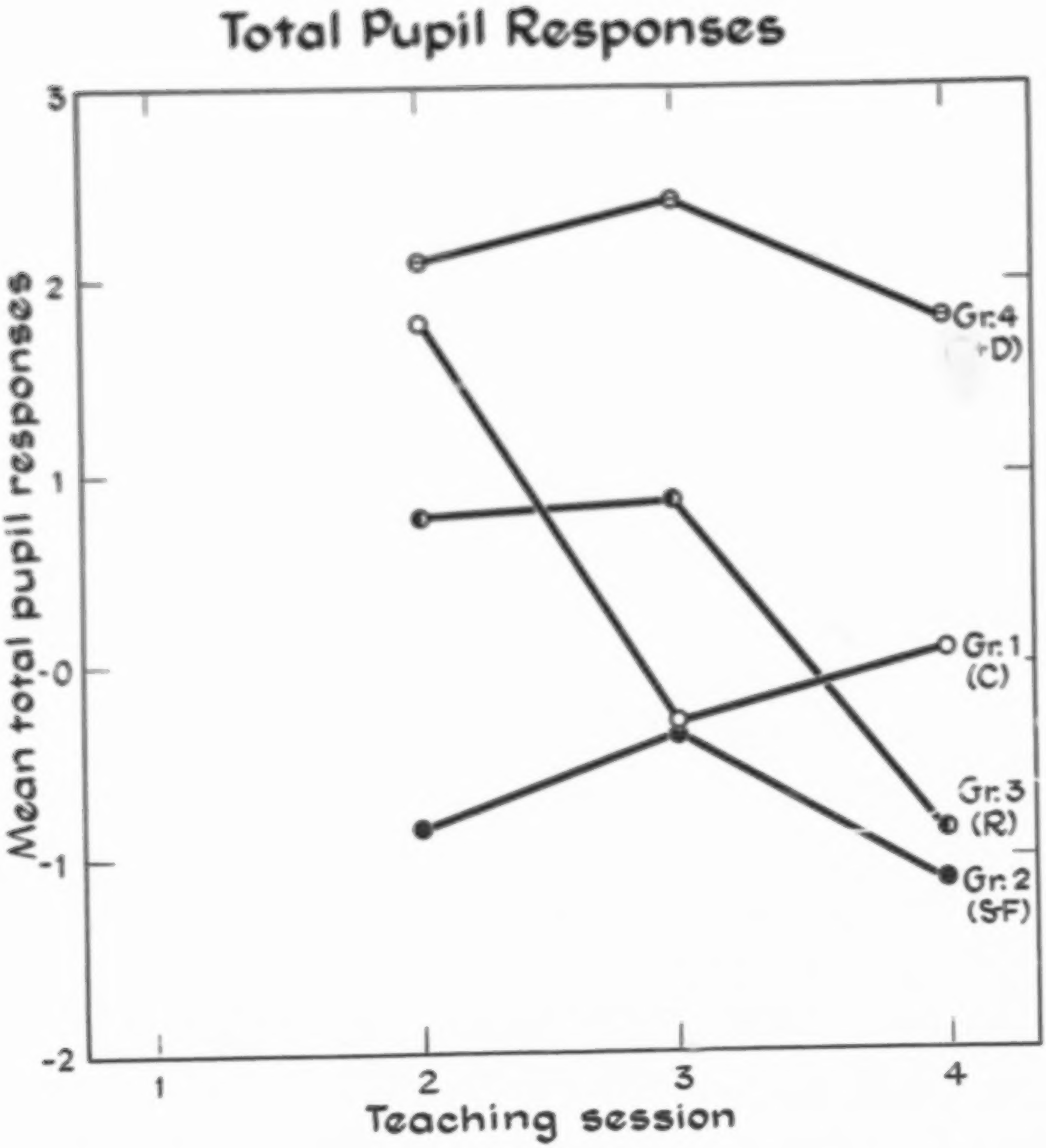

Figure 4, Experiment I

Changes in Mean Number of Total Pupil Responses
for Each Experimental Group.
(Scores adjusted to a mean of zero)

reinforcement from trial 1 to trial 4. Note that the control group subjects tended to increase their negative verbal reinforcement rate throughout treatment.

Pupil Responses: What are the probable effects of the above types of teacher training on pupil behavior? To answer this question, total pupil responses and relevant component responses were analyzed. A summary of the analyses of covariance (Table 8) performed on total pupil responses with trial 1 scores as covariates, shows that shifts in pupil responses closely followed concomitant shifts in teacher positive reinforcement. (See Table 8, Experiment I) while the increase in the Group (R) pupil sample appears short-lived, Group (R + D) pupils maintain significantly higher response levels in trial 4 (p=.005) as well as for trials 2 (p=.001) and 3 (p=.01). These data are illustrated in Figure 4. (See Figure 4, Experiment I) Note that while the control group initially showed a higher mean frequency of responses it dropped slightly over four trials. Group (R) and (R + D) pupils increased from trial 1 to trial 4.

The increase in total pupil responses immediately leads one to ask whether they are due to increased teacher positive reinforcement, or perhaps more simply, to increased questioning by the teacher.

While it is clear that differential feedback and reinforcement affected teacher behavior, it does not necessarily follow that increased positive reinforcement as defined here had an effect on pupil behavior. A comparison of volunteered pupil responses as opposed to teacher-solicited pupil responses is relevant. If increased pupil responses were largely

TABLE 8 (EXPERIMENT I)

SUMMARY OF THE ANALYSES OF COVARIANCE FOR THE EXPERIMENTAL GROUPS, WITH TRIAL ONE SCORES AS COVARIANTS AND TOTAL STUDENT RESPONSES AS THE DEPENDENT VARIABLE

	Group 1 (C)	Group 2 (S-F)	Group 3 (R)	Group 4 (R+D)	df	F
	Adjusted Means and Standard Errors					
Trial 2	72.59 5.5354	47.39 5.2296	65.93 5.3014	83.29 5.1161	3/60	8.267***
Trial 3	58.97 6.6169	54.26 6.0530	60.06 6.3019	84.31 5.9179	3/62	5.082**
Trial 4	56.92 6.604	44.94 6.517	47.82 7.752	74.12 6.309	3/52	4.004*

*	.05 level of significance
**	.01 level of significance
***	.001 level of significance

due to increased questioning by the teacher then one would detect an increase in directly solicited answers, and a decrease or no change in pupil-volunteered statements and questions. This does not appear to be the case. F ratios based on covariance analyses of Pupil-Volunteered Statements (Table 9, Experiment I) were significant for trials 2 (p=.01) and 4 (p=.01). Group (R + D) subjects after showing sharp gains dropped off somewhat in trial 3, but rose again in trial 4. Note however, that control group in frequencies followed a similar pattern while those for groups (R) and (S-F) moved in opposite directions. Figure 5 illustrates these trends. (See Figure 5, Experiment I).

Figure 6 illustrates the relationships between teacher-specified responses and pupil responses in relation to all pupil responses. It is a graphical representation of the relevant correlation coefficients reported in the next section. Mean treatment frequencies of those responses directly solicited by the teacher (DSI and DSG-1) and pupil-volunteered statements and questions (V and V?) were summed for each trial. This sum was then divided by total pupil-volunteered responses (V + V?) so that the proportion of pupil-volunteered responses to teacher-solicited responses could be determined for each trial.[1] Note that from trials 1 to 4, an increasing percentage of volunteered responses contributes to the combined

---

[1] $\frac{(V + V?)}{(DSI + DSG\text{-}1) + (V + V?)}$

TABLE 9 (EXPERIMENT I)

SUMMARY OF THE ANALYSES OF COVARIANCE FOR THE EXPERIMENTAL GROUPS, WITH TRIAL ONE SCORES AS COVARIANTS AND TOTAL STUDENT VOLUNTARY STATEMENTS AS THE DEPENDENT VARIABLE

	Group 1 (C)	Group 2 (S-F)	Group 3 (R)	Group 4 (R+D)	df	F
	Adjusted Means and Standard Errors					
Trial 2	13.23 2.4159	5.98 2.3169	12.51 2.2934	17.31 2.2422	3/60	4.184**
Trial 3	10.14 2.3562	8.20 2.1965	10.95 2.1850	15.10 2.1588	3/59	1.786[a]
Trial 4	12.27 2.3098	7.30 2.3056	8.80 2.5336	17.41 2.0725	3/48	4.159**

[a] 250 level of significance

** .01 level of significance

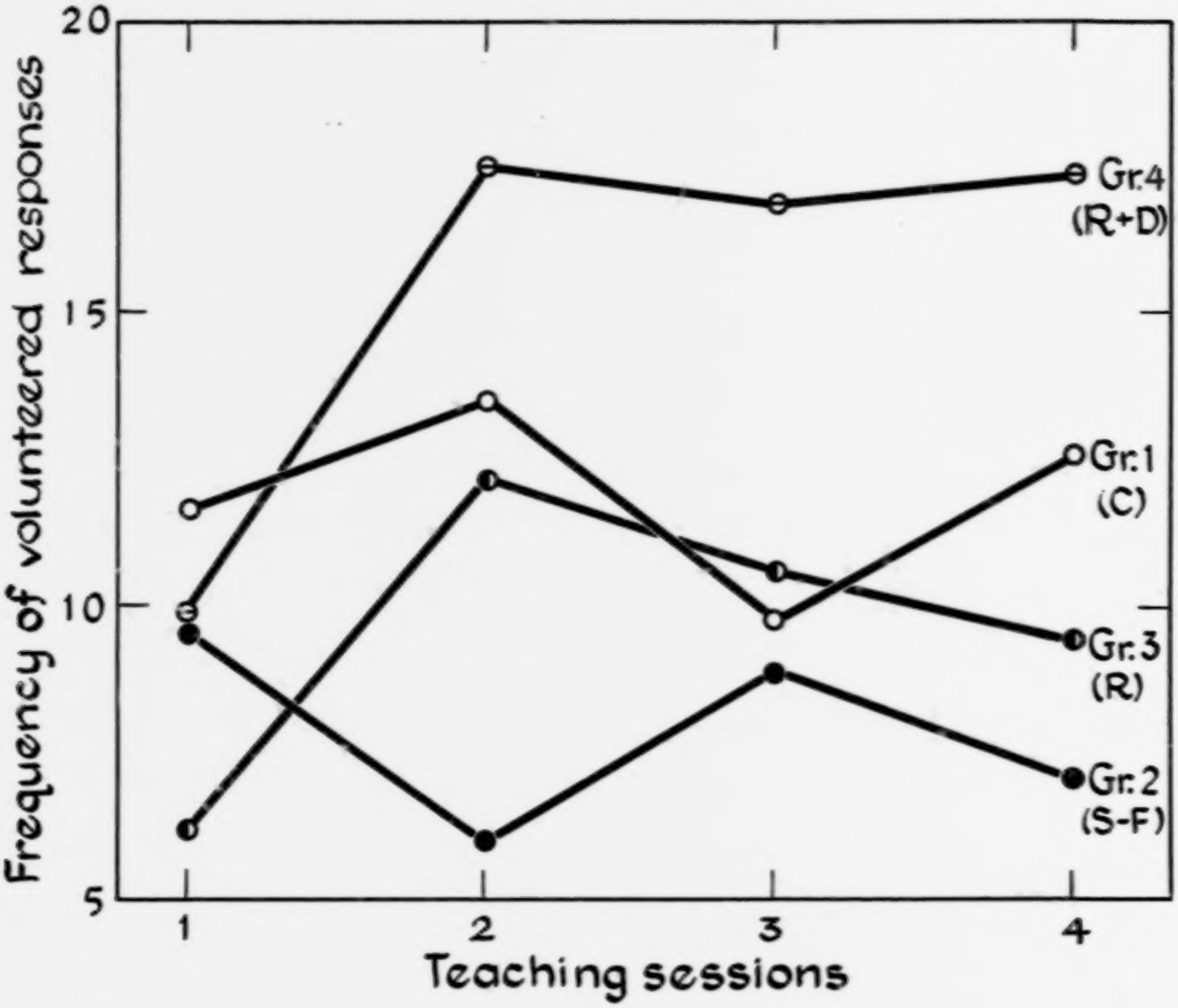

Figure 5, Experiment I

Changes in Frequency of Pupil Volunteered Responses Across Teaching Sessions for Each Experimental Group.

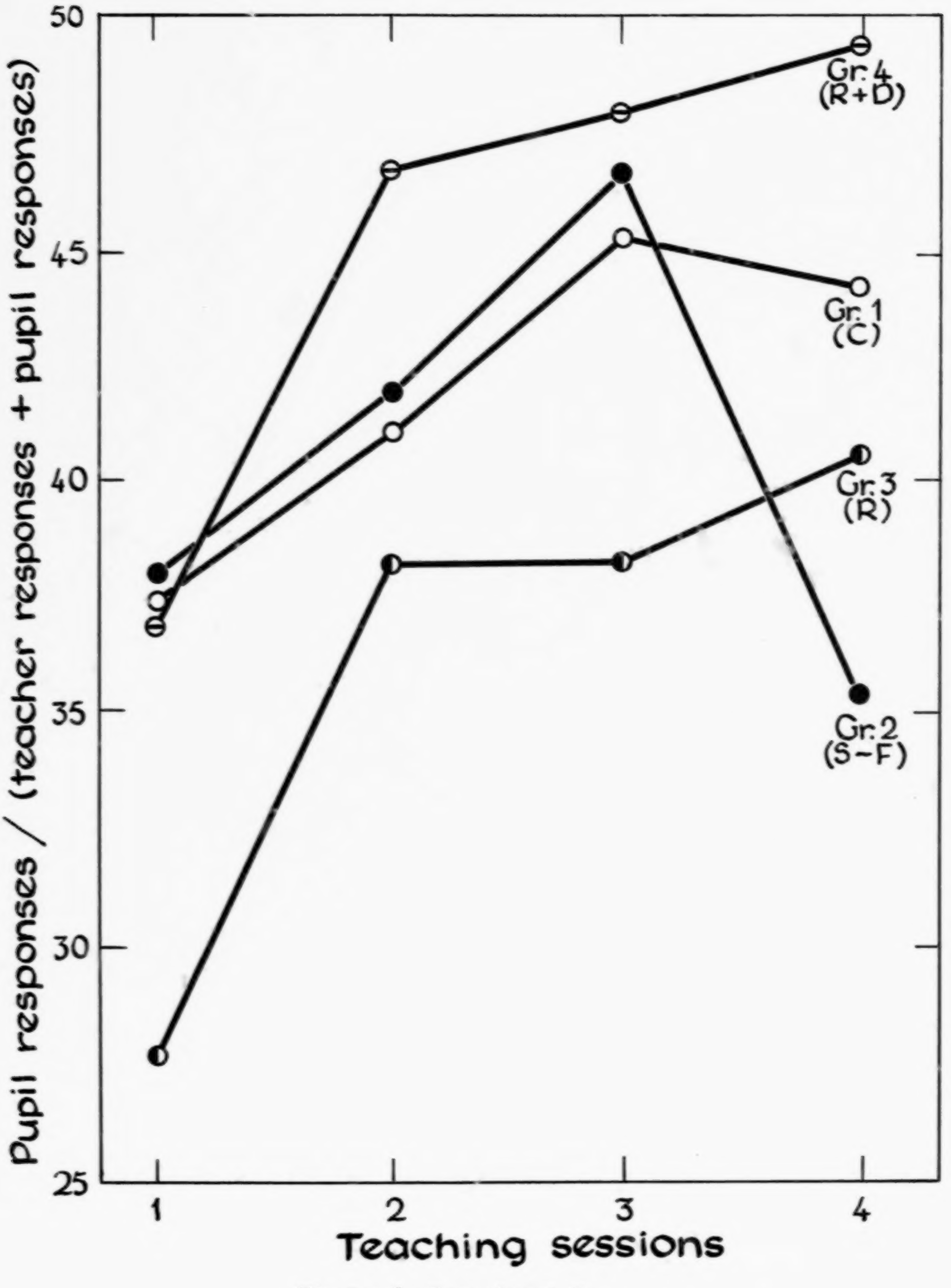

Figure 6, Experiment I

Proportion of Pupil Initiated Response in Relation to Combined Teacher and Pupil Initiated Responses for Each Experimental Group Across Teaching Sessions.

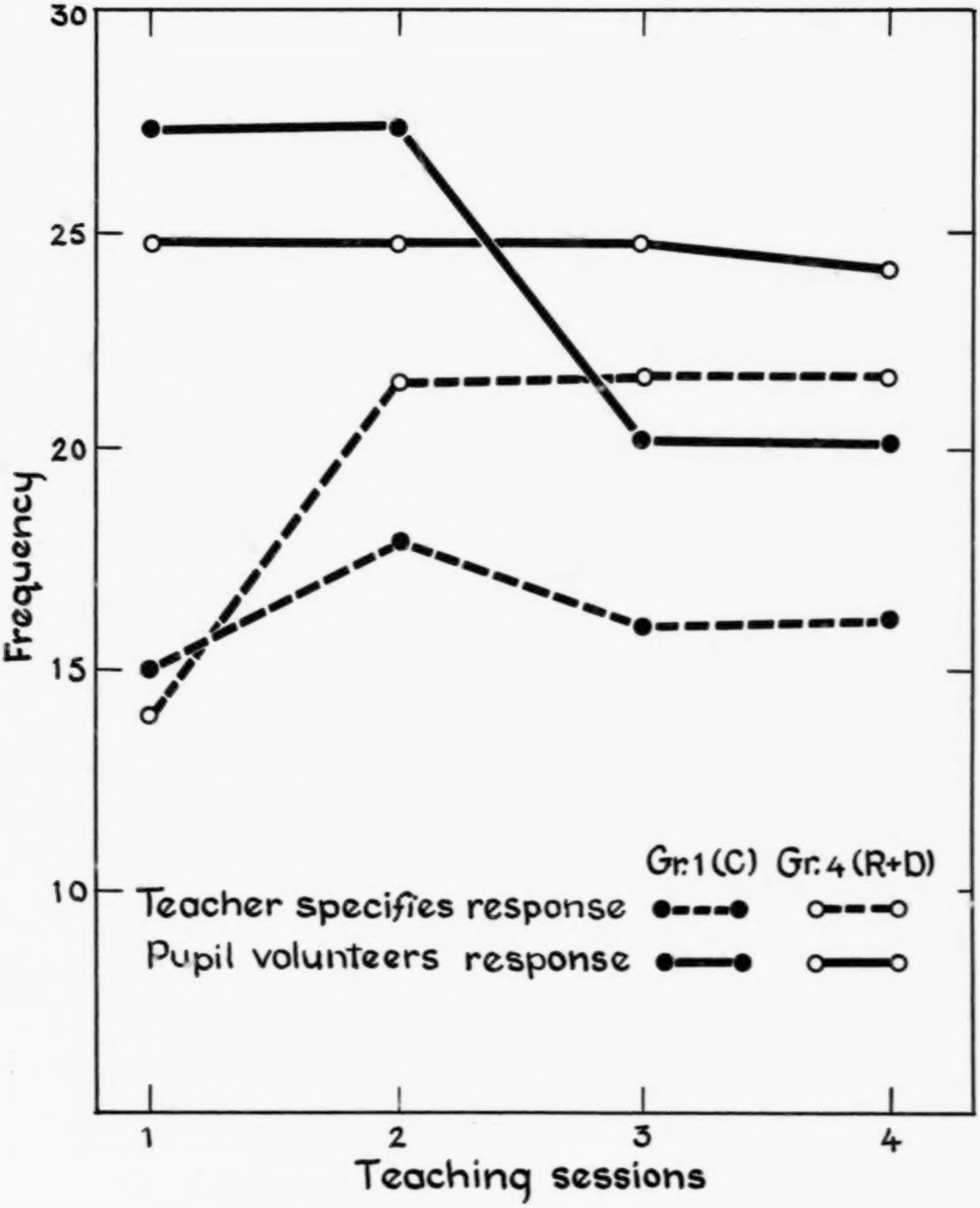

Figure 7, Experiment I

Adjusted Mean Frequencies of Teacher-Specified and Pupil-Volunteered Responses for Maximum Treatment (R+D) and Control Groups (C).

totals in Groups (R) and (R + D) while the control group shows smaller increases. Self-feedback subjects show a distinctly different pattern. This decrement is consistent with their performance in terms of teacher reinforcement.

In Figure 7, teacher-specified (DSI + DSG-1) and pupil-volunteered responses (V + V?) are presented for the controls and Group (R + D). The frequencies plotted are based on unadjusted treatment means.

Finally, it is of interest to consider intercorrelations between various response categories of the dependent variable, and total pupil responses. The intercorrelation matrix for these relationships is presented in Table 10. (See Table 10, Experiment I). These correlational data are consistent with the earlier reported results. Teacher positive reinforcement was found to be significantly related to total pupil responses (.50) and to volunteered pupil statements (0.46). Teacher-specified questions were also significantly related to total pupil responses (0.56). However, when they are considered in relation to volunteered pupil responses, the relationship is very slight (+0.07), and negative (-0.13) when we consider volunteered questions in relation to teacher-specified questions which are directed to the class as a whole.

An initially surprising relationship obtains between total pupil responses and negative verbal reinforcement by the teacher (+0.45). One interpretation would be that negative verbal reinforcement produces pupil attention. In addition, it also probably has feedback-value--it will be recalled that "No" and "Wrong" responses by the teacher were included in the negative reinforcement response category.

TABLE 10
(Experiment I)

INTERCORRELATION MATRIX FOR THE MAJOR RESPONSE CATEGORIES OF THE DEPENDENT VARIABLE

Response Category		TPR	DSI	DSG-I	DSG-V	V	+VR	+NVR	-VR	-NVR
		1	2	3	4	5	6	7	8	9
Total Pupil Responses (TPR)	1	1.00	0.56	0.09	0.41	0.56	0.50	0.16	0.45	0.07
Teacher Question to Individual Pupils (DSI)	2		1.00	0.08	0.13	0.07	0.33	0.03	0.22	0.02
Questions to Class, then to one Individual (DSG - I)	3			1.00	0.02	-0.13	0.09	0.02	0.02	-0.03
Question to Class, then a Pupil Volunteer (DSG-V)	4				1.00	0.14	0.24	0.17	0.11	0.10
Volunteer Statement by Pupils (V)	5					1.00	0.46	0.23	0.20	0.007
Teacher Positive Verbal Reinforcement (+VR)	6						1.00	0.44	0.14	0.007
Teacher Positive Non-Verbal Reinforcement (+NVR)	7							1.00	0.02	0.14
Teacher Negative Reinforcement (-VR)	8								1.00	0.15
Teacher Negative Non-Verbal Reinforcement (-NVR)	9									1.00

Discussion of Results

It is clear that feedback conditions proved to be the most effective training arrangement. Perhaps of greatest interest are what appear to be those variations in feedback which are most effective. Clearly, adding cue discrimination to the training method substantially improves the procedure. This procedure, however, is the most "costly" in that it requires the active involvement of the experimenter to describe salient cues and to suggest ways of reinforcing participating behavior that the subject could use.

Variations in the effectiveness of this procedure might occur when a variety of experimenters are used. Informal observations suggest that not all trainees responded equally positively to this condition. Trainee characteristics probably interact with experimenter characteristics, and such interactions probably influence differentially the effects of feedback and cue-discrimination training.

Equally interesting is the relative ineffectiveness of the self-feedback condition, attractive because it is the least costly procedure. This method is probably ineffective because the desired response is not adequately cued. Even if trainees had a limited response repertoire of reinforcing responses, they still could have used them consistently, if somewhat monotonously. But, the rate increase is not likely to occur if the trainee does not "know" when to emit the desired response. Both of the other feedback conditions cue as to appropriate response in some form.

This self-feedback condition might be improved by introducing some cueing procedures. Or, a combination of viewing models and self might be effective. The results obtained in this experiment indicate only that a limited kind of self-viewing, presumed to be a self-feedback condition and designed to be so, is not highly effective in producing behavior change.

The results of this experiment suggest that the operant conditioning model may be extended to situations in which the learner is not actually behaving but merely watching his performance after the actions have occurred. This extrapolation, if further substantiated. greatly increases the application of this particular paradigm. However, further research must also be directed to an analysis of the viewing conditions--characteristics of the persons viewing, time interval between enacting and viewing, the kind of behavior being reinforced, and similar conditions which might reasonably be expected to enhance or to limit the effects of the reinforcement procedure.

Also of theoretical interest is the possibility of modifying the operant conditioning paradigm by instructing subjects. This procedure shortens the time and cost of shaping the desired behavior through a series of successive approximations. The learning paradigm of this arrangement needs explication and analysis.

CHAPTER III

[redacted]

EXPERIMENT II

EFFECTS OF FEEDBACK AND PRACTICE CONDITIONS ON THE ACQUISITION OF A TEACHING STRATEGY

Material for this study may be found in Appendices G and H.

A key problem in the development of instructional systems designed to produce teaching skills is that of providing adequate feedback on the teacher's performance. Quite apart from the practical limitations encountered in trying to schedule immediate feedback sessions with teacher trainees, supervision tends to focus on the end products of performance rather than the course of learning. This is simply due to the fact that it is not usually possible to provide feedback during the teaching performance. Discrimination training is necessarily based on the perceptions of the teacher and the supervisor as they remember the lesson. The entire process thus tends to invite a heavy reliance on private frames of reference. The supervisor and the teacher do not start from a common perception of what was done, how it was done, and what the effects were.

Perhaps the signal advantage of televising trainee lessons is that in subsequent supervision sessions, the original performance can be completely reinstated. In this way, the teacher intern is not required to respond to supervision on the basis of what he and the supervisor recall about a complex series of events. Further, the usual inhibitory effects of delayed feedback may be offset, not only because the initial teaching performance is reinstated, but because it is possible to provide discrimination training at any point in the development of a response sequence. Rather than reacting to performance in terms of end products, it is possible for the intern and the supervisor to analyze a given interaction sequence in the classroom; identify salient cues; develop strategies to improve further performance. The results of the experiment reported in the preceding chapter support these hypotheses. Discrimination training during videotaped playbacks of intern lessons produced significantly greater increases in selected teacher behaviors than did confirmation or self-feedback procedures.

If the above reasoning is valid, the reinstatement of original performance by videotape does in fact solve the problem produced by delaying feedback and interspersed practice sessions take on new meaning. For example, given the capability to reinstate the original performance by videotape at any point in time, should practice precede or follow supervision? In the first case, the acquisition of a complex skill might be slower. However, substantial gains might be realized because the learner becomes less defensive about feedback based on his earlier performance, and thus becomes more responsive to supervision. This would be "delayed" feedback in the sense that the learner was receiving training based on an earlier performance. In this kind of a training situation, the learner would be serving as his own model.

Where supervision immediately followed performance, the learner's ego-involvement in the lesson just taught might lead him to be less open to suggested change. It is also probable, of course, that the immediacy of supervision would outweigh such potential resistance.

There is a general consensus that initially at least, massed practice is optimal for the acquisition of complex skills and reasoning strategies. Given time limitations in training, however, we are still left with the problem of retention. In the study described in the preceding chapter, it was found that while interns quickly reached criterion (where the dependent variable was amount of reinforcing of pupil participation) initial gains were not lasting. This result suggests that distributed practice would be a logical alternative. Further, given the perceptual adequacy of videotaped feedback, it is quite possible that acquisition rates would not be significantly lower than under massed practice conditions.

Finally, while there is considerable evidence to suggest that distributed practice is generally superior to massed practice, why this is so is not known. However, as Hilgard (1962) points out, none of the current theories account for exceptions to the general rule. In training, then, the optimal organization of practice periods becomes an inelegant process of empirically establishing the limits of each approach in terms of a given skill or strategy. The purpose of the experiment reported here was to compare several methods of distributing practice and feedback when the latter employed videotaped performances of the learner.

General Procedure: The general procedures followed in this research were similar to those reported in the first experiment. Intern teachers were videotaped on four separate occasions, during the first 20 minutes of regular classroom lessons. In the intervals between each of the taping sessions, they received experimental treatments within the context of regular supervision.

Supervisory sessions were alike for all subjects who viewed videotaped playbacks of their earlier teaching performance with an experimenter who provided discrimination training. However, instead of varying the amount of feedback within each session as in the first experiment, we held within-session feedback constant, and manipulated the amount of practice and delayed feedback over four experimental groups. In addition, the posttest was videotaped approximately seven weeks after training so as to permit inferences about retention as well as acquisition curves.

The Dependent Variable: In the earlier experiment, interns had been trained to reinforce pupil participation in discussion and review lessons. A natural extension of this skill seemed to be those techniques where the teacher could increase the quality of such participation.

Rather than attempting to operationally define a "penetrating question", or to develop techniques for suppressing superficial "first-answer" pupil responses, the approach taken was to develop classroom techniques which followed simple shaping procedures. Interns were given three basic rules: 1) Do not give immediate answers to pupil questions. 2) Once a pupil has responded, try to get him to "go beyond" the information given (by one of the several specified techniques). 3) Differentially reinforce pupil responses that demonstrate increased critical awareness.

In the initial written instructions to each intern, and in supervision as well, a series of discrete techniques were presented following the statement of each rule. The basic strategy in each case was the same. Following a pupil response, the teacher asked a question designed to elicit more information or more meaning from the pupil. If the pupil response was adequate or insightful, the teacher then attempted to get him to relate the questions, answer, or comment to another area, or to spell out its implications for a given issue.

These various techniques were presented as exemplars of a basic questioning strategy termed probing. Specific techniques were defined as: clarification, critical awareness, redirection, prompting and refocus. (The term used to describe each specific category reflects the teacher's goal when using a given technique).

In addition, a secondary set of techniques designed to achieve the same ends as probing procedures were included as part of the training problem.

The latter techniques differ from probing in that they may or may not depend on a prior pupil response, i.e., the teacher can introduce them at any point in the discussion himself, or use them as specific types of probes. These techniques were termed encouraging alternatives (divergent thinking), supposition (role play in brief) and pupil summary of discussion. Both in training and in later tape analysis, the pupil response served as an S^d to cue the intern or rater that a teacher probing (classifiable) response should immediately follow.

Treatment Conditions: Stanford Interns were assigned to one of four experimental groups, each of which received varying amounts of practice between feedback sessions. Subjects in each group received discrimination training from one of two E's who were present at each feedback session. All feedback sessions were thirty minutes in length. A summary of the design appears in Table 1. (See Table I, Experiment II).

TABLE 1 (EXPERIMENT II)

SUMMARY OF THE EXPERIMENTAL DESIGN

Practice Conditions	Feedback on Teaching Performance	
	Immediate	Delayed
Massed Practice	Group 1 (N = 21)	---------
Distributed Practice (1 week intervals)	Group 2 (N = 21)	Group 3 (N = 21)
Distributed Practice (2 week intervals)	----------	Group 4 (N = 22)

TABLE 2 (EXPERIMENT II)

VIDEOTAPING AND FEEDBACK SCHEDULE

		Pretest	Trials		Posttest	
		Day 1	Day 2	Day 3	Day 4	
GROUP I	Original Tape Feedback Session	1 1	2 2	3 3	45 (45)	Massed Practice Immediate Feedback
GROUP II	Original Tape Feedback Session	1 1	7 7	14 14	45 (45)	Distributed Practice Immediate Feedback
GROUP III	Original Tape Feedback Session	1 7	7 14	14 21	45 (45)	Distributed Practice Delayed Feedback
GROUP IV	Original Tape Feedback Session	1 7	14 21	28 35	45 (45)	Distributed Practice Delayed Feedback

Group 1. (Immediate Feedback -- Massed Practice): Group 1 subjects, like those in the other three groups, initially received written instructions on probing. The instructions related the new techniques to the interns' prior training in reinforcement, and stated that the basic aim of the current study was to help interns develop a broader range of basic questioning skills in discussion and review lessons. Each rule was then stated, and specific techniques together with relevant examples were presented. One of the two E's then briefly reviewed the techniques with the intern, reinforced positive statements about the potential utility of probing techniques, and viewed intern's playback with him.

Each time a pupil responded verbally, E cued the intern, and if the latter had probed, then E reinforced him. If he had not, E suggested how this might be done. Interns in the first group were both videotaped at the school on a given day, and received supervision based on that performance the same evening. The posttest was videotaped 45 days after the pretest was made. (See Table 2 for videotaping schedule.)

Group 2 (Immediate Feedback - Distributed Practice): These subjects received the same kind of discrimination training as those in Group 1 . Instead of receiving massed practice however, they were directed to practice probing techniques for a one-week period following each of the supervision sessions.

Supervision sessions were based on videotapes that had been recorded earlier on the same day. The posttest followed the pretest videotape by 45 days.

Group 3 (Delayed Feedback -- Distributed Practice): Subjects in this group received the same supervision treatment as those in Group 2. However, they always viewed a videotape of their performance which had been filmed one week earlier. Unlike Group 1 and 2 subjects who received feedback based on that day's performance, these subjects were taped on the day following supervision, and received no discrimination training based on their performance for one week. At the conclusion of each session, Group 3 subjects were directed to practice probing techniques not only in the next day's lesson but during the rest of the week as well.

Group 4 (Reinstated Feedback -- Distributed Practice:) These subjects received treatment distributed over a six-week period. Following taping on the first day of week one, they received supervision based on that tape on the first day of week two. They were taped and supervised on alternate weeks. This group was also posttested 45 days after the pretest.

Subjects: Interns were selected from the same sample as that for the first experiment.[1] A total of 85 intern teachers majoring in English, social studies, mathematics or science were selected for study. Approximately equal numbers of interns from each subject-matter area, and from socio-economically equivalent schools were assigned to each of the four groups. Mean age for the four groups varied from 23.7 to 35.5 years. Sex differences and subject-matter in each group were very similar to those already described in the first experiment.

---

[1] There was a three-month break between the conclusion of the first study and the beginning of the study being described here.

Measurement Procedures: Four videotapes for each intern in each group were analyzed by raters trained for this purpose. Four raters received intensive training on non-experimental tapes before the analysis of the latter was begun. Training consisted of joint rating sessions that continued until there was 95% agreement on the major response categories.

Reliability was maintained by frequent analyses of double-rated tapes. In addition, joint-rating sessions were held after each block of 50 tapes had been completed so that systematic rater biases could be controlled. As rating progressed, it was found that certain tapes inevitably yielded low interrater agreement. This occurred when the intern in question was unable to maintain classroom discipline, or when there was an unusually high rate of interaction combined with "fuzzy" audio or visual output. In these cases, the tapes were rated by independent teams of raters -- by dividing the rating task between themselves and replaying frequently, two raters were able to record all interactions. To control for combined rating effects, independent teams were used. The identification of these "trouble" tapes posed a problem. When the raters were asked to rank the rating difficulty of tapes, their judgments were not entirely consistent with reliability checks. For this reason, it was finally decided to double-rate all tapes.

The results of these control procedures are reflected in the reliability coefficients reported in Table 3. (See Table 3, Experiment II). The videotapes on which they are based were selected in the following manner. Since coefficients based on all of the tapes would have been prohibitively expensive and also time-consuming, 160 tapes were selected

TABLE 3 (EXPERIMENT II)

RELIABILITY COEFFICIENTS BASED ON A RANDOMIZED SAMPLE OF 80 VIDEOTAPES RATED BY TWO INDEPENDEND OBSERVERS

RESPONSE CATEGORY	RELIABILITY COEFFICIENT
Total Pupil Responses	0.9934
Total Teacher Probes	0.9939
Total Teacher Non Probes	0.9871
Total Teacher Reinforcement	0.9579
Probing Sub-Categories:	
a. Clarification	0.9689
b. Critical Awareness	0.8658
c. Redirect	0.7633
d. Prompting	0.5914
e. Refocus	0.7312

using tables of random numbers. Five tapes from each group for each trial were randomly selected from pools of 15 to 21 tapes.[1] The coefficients reported then are based on a representative selection of independently rated tapes from each group and for each trial. As can be seen interrater agreement on the major response categories is very high.

A certain amount of data were lost between initial videotaping and statistical analysis. Forty-four tapes were omitted at the outset because of technical inadequacy or unduly short tape-time. Fifty-six tapes were prorated to bring them up to the 20.0 minute criterion. Tapes less than 15 minutes in length were omitted; those between 15 and 20 minutes were prorated. Before any of the above adjustments were made, the T statistic was applied to mean tape-time and number of omitted tapes per cell to determine whether or not any significant differences between cells existed. The results were well short of significance as these tapes were almost equally divided among all cells. The standard errors for short tapes varied from .40 to 1.2 minutes. As a further check on the data, the mean number of days between pre- and post-test were calculated for each experimental group. The overall mean was 45.25 days, with $S_e = 1.41$ days.

RESULTS

Probing was presented as a basic questioning strategy in training, and specific techniques such as Refocus or Critical Awareness were subsumed under this broad general concept. It is important to note that this was a logical distinction made to facilitate training. It does not follow that it is psychometrically relevant. In fact, the analysis of results did not proceed in terms of one broad dependent variable, but in terms of eight. Each of the questioning and related techniques were viewed as discrete dependent variables. Table 4 provides support for this procedure. The intercorrelations among the subcategories of probing are low. Of the 15 intercorrelations, 5 are significantly different from zero. Of these, 3 reach the .05 level and 2 reach the .01 level. Prompting is significantly related to redirect ($r = .24$, $p = .05$), refocus ($r = .28$, $p < .05$) and clarification ($r = .36$, $p < .01$). However, prompting occurred infrequently (group means varied from 3.7 to 1.7) and contributed little to overall differences. Each of the subcategories correlates significantly with probes.

---

[1] Team ratings were not included in selection because they had been rated by special techniques, and were known to be highly reliable on all major response categories.

In the analysis of treatment differences then, F ratios for all of the major probing techniques must be considered. F ratios based on probing provide a general overview of the results taken as a whole, but since they include variables which did not reflect significant differences, there is a suppression effect.

Treatment Differences: Analysis of covariance with relevant trial 1 scores as covariates were carried out to determine between-group differences in trials 2, 3, and 4. Table 5 summarizes these results. As can be seen, differences are reflected in several of the dependent variables.[1]

---

[1] In general, the most frequently occurring responses are listed first in the tables. Further, the more frequently a response occurs, the more highly correlated it is with probing.

TABLE 4 (EXPERIMENT II)

CORRELATION MATRIX FOR THE MAJOR RESPONSE CATEGORIES OF THE DEPENDENT VARIABLE

N = 79

Dependent Variables	NPr	AQ	TNR	Pr	Cl	CA	Rd	Pmt	Rf	EA	PS	RP
	19	20	21	22	23	24	25	26	27	28	29	30
No Probes (NPr)	1.00	.4784	.6889	.0960	.0542	-.0246	.1925	.1265	-.0012	.0554	.3089	.1860
Answers Questions (AQ)		1.00	.4726	-.0803	-.1199	-.0599	-.0046	.1614	-.0333	-.0697	.1673	.1084
T. No Response (TNR)			1.00	.1780	.1197	.0399	.0563	.2121	.1070	.0678	.1576	-.0000
Probes (Pr)				1.00	.9121	.5072	.5068	.4557	.4027	.1099	.0380	.0858
Clarifies (Cl)					1.00	.2385	.3269	.3697	.2126	.0233	.0251	.0476
Critical Awareness (CA)						1.00	.1796	.1728	.1338	.0584	.0591	.1313
Redirect (Rd)							1.00	.2416	.2049	.1096	.1351	.0292
Prompt (Pmt)								1.00	.2834	-.035	-.0016	.0491
Refocus (Rf)									1.00	.0349	.1223	.0855
Encouraging Alternatives (EA)										1.00	.0214	.1569
Summary (PS)											1.00	.0212
Role Play (RP)												1.00

NOTE: In order for any of the coefficients reported above to be considered significantly different from 0, r must = .232 (p = .05) or .303 (p = .01).

TABLE 5 (EXPERIMENT II)

SUMMARY OF F RATIOS AND SIGNIFICANCE LEVELS FOR THE DEPENDENT VARIABLES, DERIVED FROM TRIAL TWO, THREE AND FOUR ANALYSES OF COVARIANCE, WITH TRIAL ONE SCORES AS COVARIANTS

Response Category	Trial 2		Trial 3		Trial 4	
	F Ratio	df	F Ratio	df	F Ratio	df
Probing	0.560	3/58	1.100	3/53	.053	3/52
Clarification	1.170	3/58	0.505	3/53	.200	3/52
Critical Awareness	0.990	3/58	1.887[a]	3/53	.280	3/52
Redirection	3.476*	3/58	2.631[a]	3/53	.068	3/52
Prompting	1.147	3/58	0.746	3/53	.620	3/52
Refocus	1.107	3/58	2.822[b]	3/53	3.134*	3/52
Encouraging Alternatives	0.370	3/58	1.743[a]	3/53	.516	3/52
Pupil Summary	1.122	3/58	.500	3/53	.792	3/52
Role Play	0.595	3/58	.457	3/53	1.218[a]	3/52
	N = 63		N = 74		N = 53	

Levels of Significance:

a for F = 1.41, p = .25

b for F = 2.18, p = .10

* for F = 2.76, p = .05

** for F = 4.13, p = .01

TABLE 6 (EXPERIMENT II)

SUMMARY OF THE ANALYSES OF COVARIANCE FOR THE EXPERIMENTAL GROUPS, WITH TRIAL ONE SCORES AS COVARIANTS AND PROBES AS THE DEPENDENT VARIABLE

		Group 1	Group 2	Group 3	Group 4	df	F
		Adjusted Means and Standard Errors					
Trial 2	M	12.68	8.34	15.07	11.54	3/58	0.56
	Se	3.94	3.61	3.95	4.56		
Trial 3	M	10.20	5.01	12.30	5.54	3/53	1.10
	Se	3.39	3.28	3.28	3.97		
Trial 4	M	7.79	7.60	6.05	7.84	3/52	.05
	Se	3.53	3.31	3.65	4.12		

Only one significant treatment difference, refocus ($p < .05$), appeared in trial 3, though some differences approached significance ($p = .25$) on the variables, critical awareness, redirection, and encouraging alternatives.[1] However, the relative frequency of occurrences for each of the response categories that determine treatment differences is important in attaching meaning to these results. For example, role play occurs infrequently. Clarification, which contributes heavily to total teacher responses, occurs frequently.

Treatment differences in trials 2 and 4 are much less general even when those approaching significance are included. In trial 1, redirection was significant ($p < .05$). The F ratio for probing is only 0.600. In trial 4 refocus shows treatment differences ($p < .05$).

An overall summary of treatment differences for probes derived from the analysis of covariance is shown in Table 6. (See Table 6, Experiment II). Adjusted means, standard errors and F ratios for mean frequencies of probes by groups over trials are presented. Figures 1 and 2 (See Figures 1 and 2, Experiment II) illustrate the general pattern of probing, and clarify specific between-group differences for a given trial. Both figures are based on adjusted means so that trial 1 levels of performance can be seen in relation to the other three trials. Table 7 shows the proportion each of the subcategories contributed to the variance in the total probing score.

---

[1] Although the .05 level of significance has been used in these studies to decide for accepting or rejecting the null hypothesis, when we discovered the multivariate nature of the dependent variable, we thought calling attention to nearly significant differences would portray the results more meaningfully.

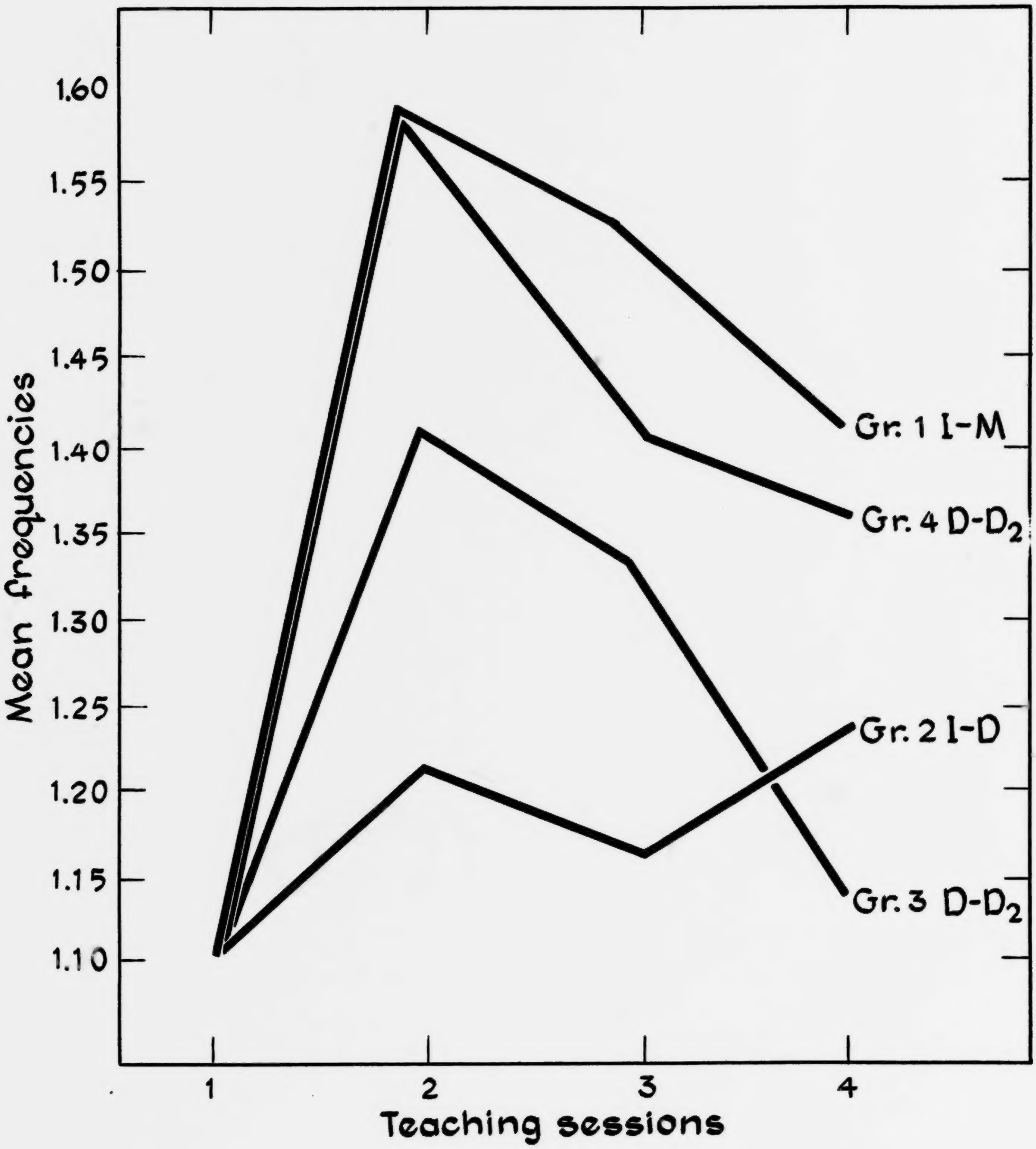

Figure 1, Experiment II

Mean Frequency of Total Probes Across Teaching Sessions for Each Experimental Group.

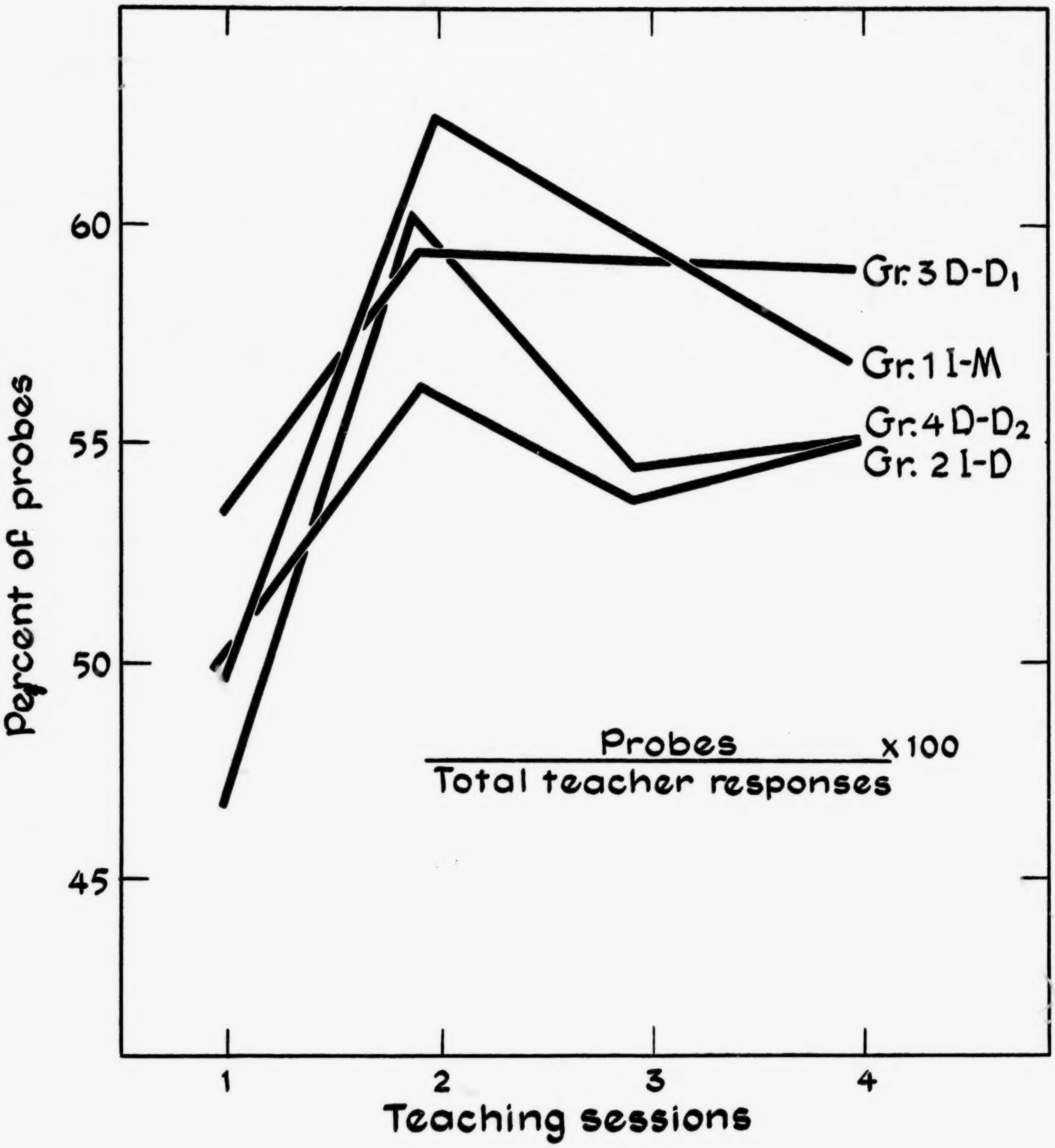

Figure 2, Experiment II

Mean Frequency of Total Probes in Proportion to Total Teacher Responses Across Teaching Sessions for Each Experimental Group.

TABLE 7 (EXPERIMENT II)

PROPORTIONAL OF VARIANCE ATTRIBUTABLE TO EACH

COMPONENT OF TOTAL PROBING SCORE

	Trial 1	Trial 2	Trial 3	Trial 4
Clarification	.28	.26	.25	.21
Critical Awareness	.18	.17	.16	.20
Redirection	.15	.17	.14	.15
Prompting	.14	.11	.13	.13
Refocus	.09	.10	.11	.10
Encouraging Alternatives	.08	.07	.08	.09
Pupil Summary	.08	.07	.07	.06
Role Play	.00	.05	.06	.06

In Figure 1, all unadjusted probing means were brought to a common point somewhat analogous to a covariance adjustment by dividing the respective means for trial one into each of the four means for each group.

Figure 1 tends to exaggerate certain group trends, particularly from trials 3 to 4. Figure 2 is singularly instructive here because it considers all of the dependent variables in relation to total teacher responses. In this figure plot points were derived by expressing probes as a percentage of total teacher responses. Now, comparing Figure 1 with Figure 2, it can be seen that Group 3 and 4 do not drop off from trial 3 to trial 4 as implied in Figure 1. Figures 3 and 4 and 5 illustrate some of the differences across trials in subcategories of probing. Recall that clarify and redirect correlate ($r = .33$), while clarify and refocus do not correlate ($r = .21$). Refocus and redirect similarly do not correlate. Because the characteristics of the dependent variable need further exploration, analyses of differences in greater detail did not seem profitable at this time. These graphs suggest that the treatments may have affected the subcategories in different ways.

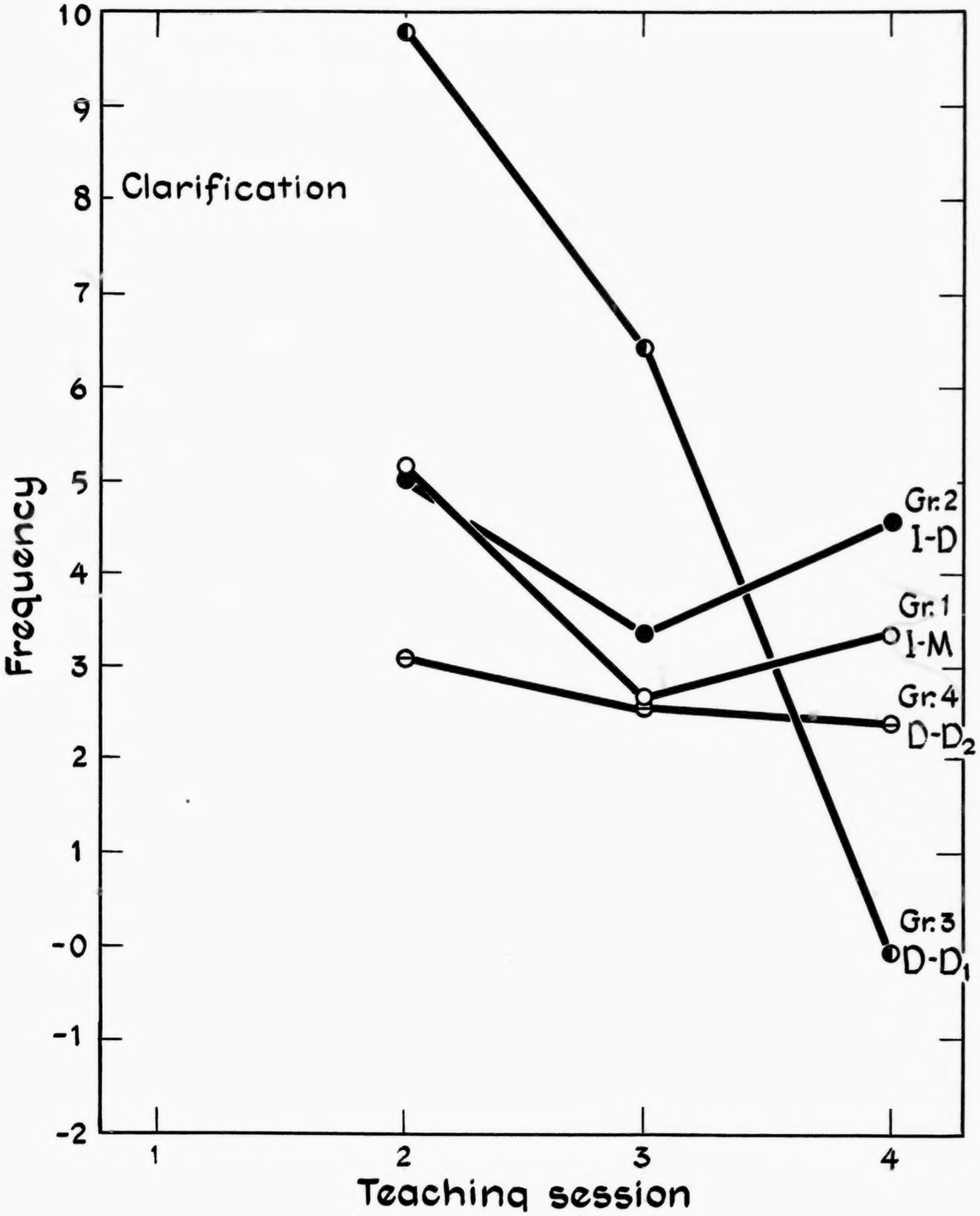

Figure 3, Experiment II

Mean Frequency of Clarification Responses Across Teaching Sessions for Each Experimental Group.

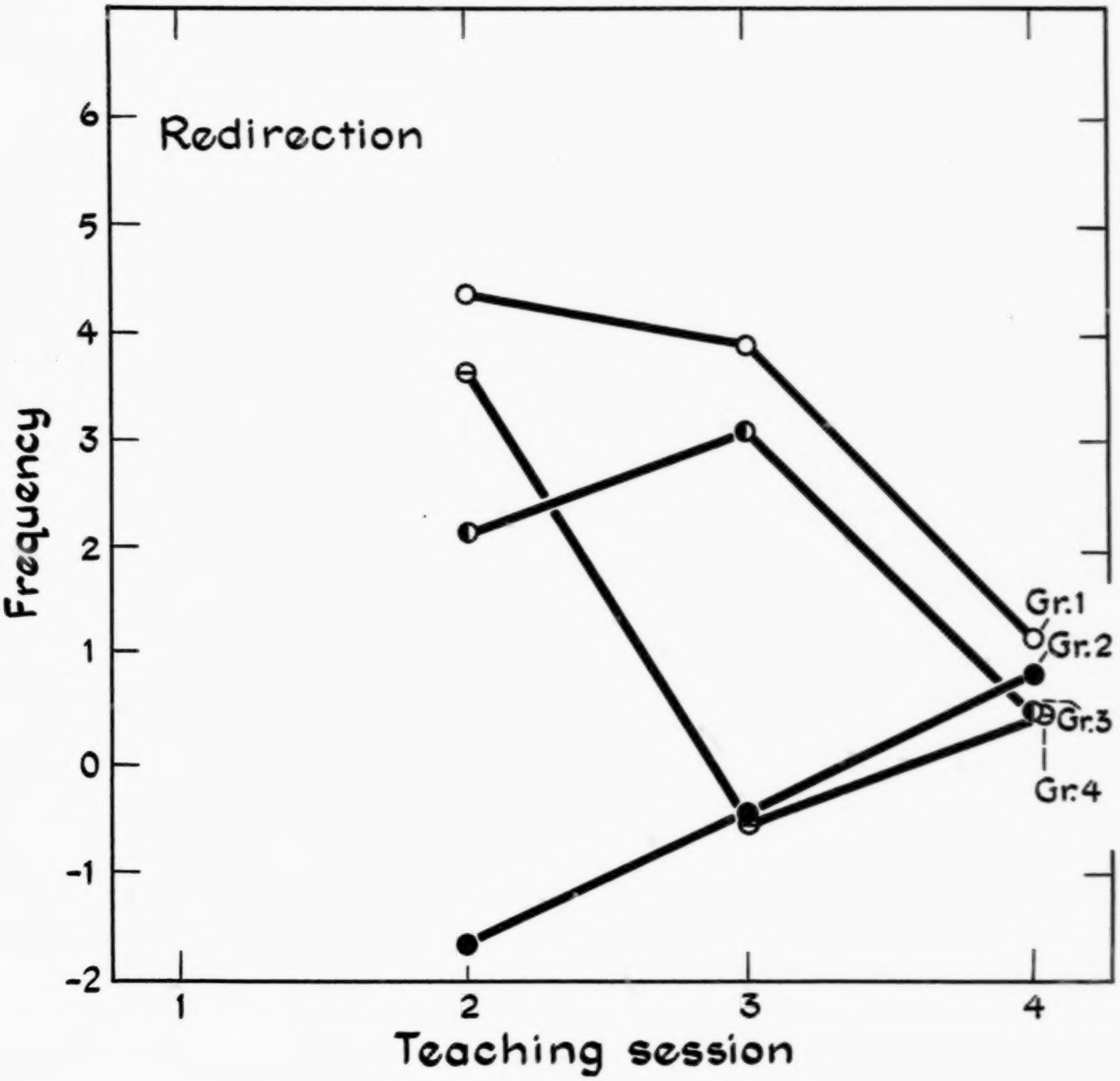

Figure 4, Experiment II

Mean Frequency of Redirection Responses Across Teaching Sessions for Each Experimental Group.

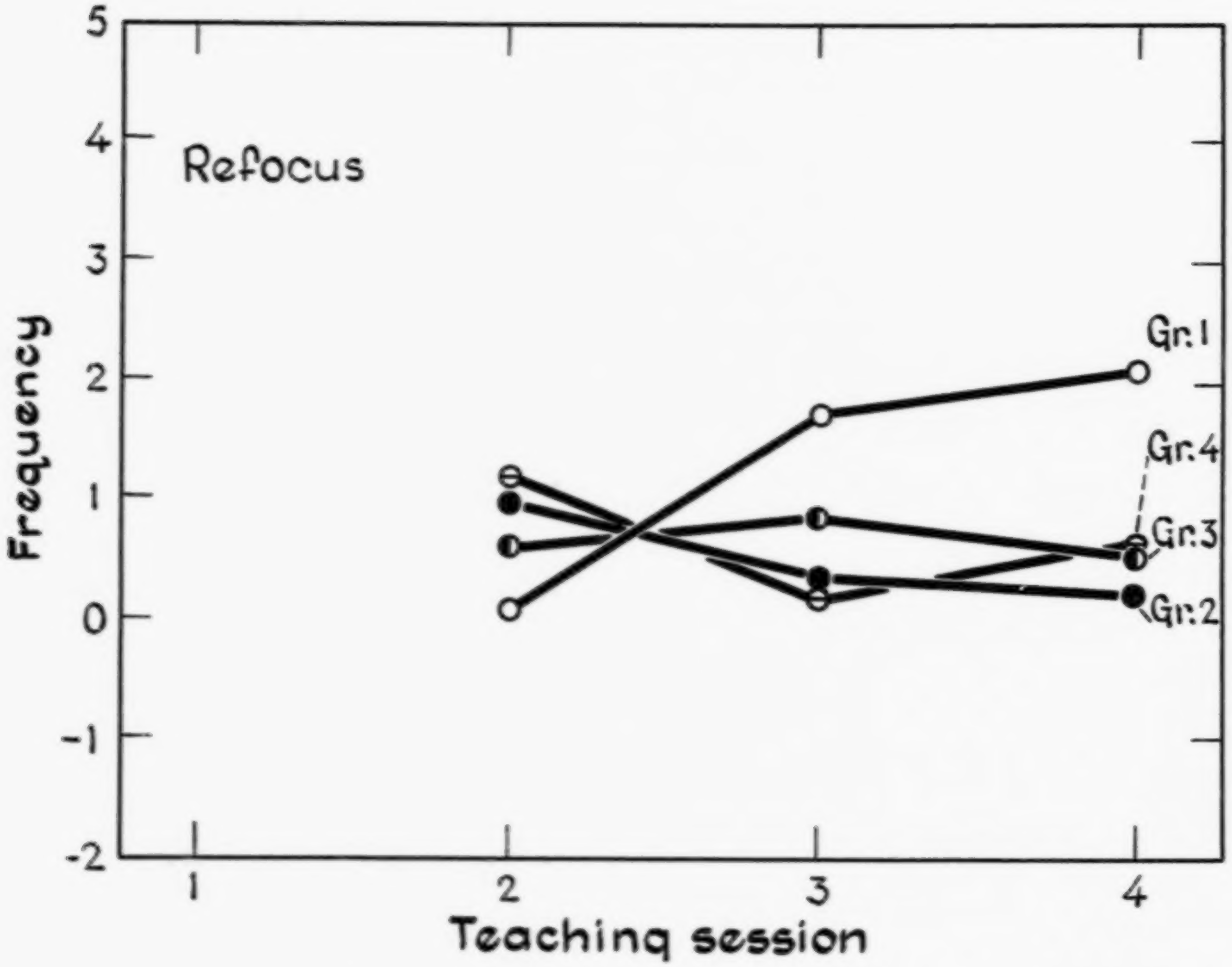

Figure 5, Experiment II

Mean Frequency of Refocus Responses Across Teaching Sessions for Each Experimental Group.

Discussion of Results

Treatment differences were not entirely consistent throughout trials and across dependent variables, so that conclusions must be qualified. Essentially no meaningful differences were found. There appears to be a differential effect of the treatments on the different categories of probing. Since an interaction analysis was not feasible, this observation must be accepted as a hypothesis, not a conclusion. Although no differences can not be treated as a conclusion, it is interesting to note that wide variations in time of feedback did not produce differences.

Certain implications can also be drawn about treatment differences related to retention. Figure 2 suggests that the distributed practice and delayed feedback groups (Groups 2 and 4) maintained relatively higher probing response rates on the posttest than did Group I who dropped off quite sharply. However, the data do not permit firm conclusions here as only one F ratio for trial 4 was significant.

Finally, it should be pointed out that there appear to be practical limits on the amount of probing possible in any given class period. Unlike teacher reinforcement which may occur in high frequency, classroom and subject matter concerns establish a ceiling for probing. The most sensitive measures of this variable are likely to be expressed as ratios or percentages related to total teacher responses.

CHAPTER IV

EXPERIMENT III

THE EFFECTS OF MODELING AND FEEDBACK VARIABLES ON THE ACQUISITION OF A COMPLEX TEACHING STRATEGY

Materials for this study may be found in Appendices G-L.

THE PROBLEM

The purpose of this study was to assess the relative effects of various arrangements of feedback and presentation variables upon the performance of a specified class of teacher behaviors.

A common approach to the transmission of teaching skills has been to provide some kind of discrimination training by means of written and oral instructions. The intern teacher typically received a description of the correct responses and their sequencing for a particular situation. He then attempts to produce these behaviors in the classroom and receives periodic feedback on his performance.

Another approach to training problems of this sort is suggested by recent findings on the role of observational learning in personality development. A review of the relevant literature by Bandura and Walters (1963b) has shown that complex social behavior may be acquired almost entirely through imitation. They state the provision of face-to-face models serves to accelerate the learning process and, in cases where errors are dangerous or costly, become an essential means of transmitting behavior patterns (Bandura and Walters, 1963b, p. 52). In addition, Bandura, Ross and Ross (1963c) have demonstrated that film-mediated models are as effective as real-life models in transmitting deviant patterns of behavior.

While experimental demonstrations of modeling effects have largely employed young children as subjects and aggression as the dependent variable, there is little doubt that adult subjects may acquire various classes of responses by the same process. The implication of this for teacher training is that an audiovisual display of a teacher model performing some desired behavior may provide an effective alternative to purely descriptive techniques.

One of the objectives of the study was to compare these two modes of presentation in the context of a televised series of training sessions for intern teachers. The experimental design permitted an analysis of the assumption that the rate and level of learning a given teaching skill varies as a function of the mode of model presentation. Two types of modeling were considered:

1. Symbolic Modeling: This is defined as a process whereby one transmits desired behaviors to the learner by means of written or verbal instructions. The subject does not view an actual portrayal of the desired behavior.
2. Perceptual Modeling: This is defined as a process whereby one transmits desired behaviors to the learner by means of a filmed model who portrays the desired behavior.

In addition to presentation variables, the problem of adequate feedback on the teacher's performance is of considerable importance in the transmission of a teaching skill. The effects of feedback have been extensively investigated in terms of reinforcement, knowledge of results, confirmation, and trial and error learning. Despite theoretical debate, there is little doubt that such feedback does produce learning (Michael and Maccoby, 1960; Hilgard, 1956).

While the experimental literature suggests that both modeling and feedback effectively change behavior, little is known about the relative effectiveness of these two processes. The human learner, because of his considerable information-processing capacities, may require little feedback when the desired behaviors are portrayed as in perceptual modeling. However, if the responses to be learned are sufficiently complex, some combination of demonstration and feedback may be required.

The purpose of the study then was twofold. In addition to permitting a comparison of symbolic and perceptual modeling, the experiment was designed to assess the relative effects of various arrangements of feedback in combination with the two presentation variables. In general, the experiment sought to determine which would be more efficient; telling the person what to do (symbolic modeling), or showing him what to do (perceptual modeling), or some combination of these approaches with feedback that includes reinforcement and further discrimination training on the relevant cues.

THE DEPENDENT VARIABLE

The dependent variable that was developed for the study is termed Probing. A complete discussion of the technique appears in Appendix G (Instructions to Interns) and Appendix H (The Rater's Manual). Thus, comments here will be brief.

Probing is a basic questioning technique in which the teacher requires students to go beyond first-answer responses. It is designed to be used in lessons where pupil participation is prerequisite to the goals of instruction, and is intended to upgrade the quality of such participation.

Once the pupil has responded by means of a question, answer or comment, the teacher may probe this response by means of one or more *Probing* techniques. These sub-classes or categories of *Probing* are termed: *Clarification*, *Increased Critical Awareness*, *Refocus*, *Prompting*, and *Redirect*. The labels in each case generally reflect the teacher's goal when using a given type of *Probing*.

Two basic distinctions have been made about *Probing* as a dependent variable in an earlier paper by Allen, McDonald, and Orme (1966). First, each of the response categories of probing acts as a discrete dependent variable. This inter-variable independence is shown in Table 1. Of all the possible intercorrelations between the various response categories, only one proved to be significantly different from zero (*Prompting* and *Clarification* $r_{P.C.}=0.31$). It follows that in the analysis for treatment effects, one can expect differential levels of significance on different variables in each session. Statistically, we are concerned with as many dependent variables as there are subcategories of *Probing*. In training, these differences were not emphasized to such a degree. Treatment was designed to produce increases in *Probing* as a general questioning strategy.

The second relevant distinction to be made about *Probing* is that there is a definite ceiling effect on its frequency of occurrence in classroom interaction. The teacher can only *probe* (or *non-probe*) following a pupil response. Since total pupil responses can reasonably be expected to vary, to some extent, independently of treatment, statistical analysis must go beyond tests of mean differences of *Probing* by group and trial. One must also consider mean *Probes* in relation to *Non-probes*. These two teacher behaviors when taken together yield a frequency which is equivalent to *Total Pupil Responses*.

TABLE 1

INTERCORRELATION MATRIX FOR THE MAJOR RESPONSE CATEGORIES MEASURED

N = 326

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
P. Questions	1	1.00	.27	.09	-.11	.12	-.10	-.09	.34	.77	.27	-.07	-.08	.05	-.12	-.06	.03	-.03
P. Comments	2		1.00	.04	.12	.53	.19	.21	.78	.22	.64	.13	.07	.20	-.09	.10	.09	-.05
P. Rhetorical	3			1.00	.01	.06	-.07	.03	.08	.25	.06	.03	-.02	.05	.00	.13	.00	-.02
P. Answers	4				1.00	.89	.51	.73	.32	.11	.22	.91	.79	.46	.37	.51	.26	.10
Total P. Resp.	5					1.00	.50	.70	.63	.09	.49	.85	.71	.49	.29	.49	.27	.06
Tchr. Reinf.	6						1.00	.46	.22	-.11	.06	.49	.50	.18	.16	.17	.13	.09
Tchr. Repeats	7							1.00	.35	-.07	.10	.65	.64	.27	.13	.36	.17	.10
Non-Probes	8								1.00	.34	.66	.12	.08	.18	-.07	.11	.01	-.06
Tch. Ans. Ques.	9									1.00	.18	-.11	-.14	.00	-.13	-.00	-.05	-.04
Tch. No Resp.	10										1.00	.18	.13	.13	-.08	.16	.15	-.02
Total Probes	11											1.00	.85	.50	.42	.56	.34	.14
Clarifys	12												1.00	.16	.14	.31	.18	.13
Crit. Aware.	13													1.00	.19	.11	.14	-.01
Redirect	14														1.00	.19	.07	.10
Prompt	15															1.00	.17	.09
Refocus	16																1.00	.04
Encg. Alt.	17																	1:00

Note: In order for any of the coefficients reported above to be considered significantly different from 0, r must = .195 ($p<.05$) or .254 ($p<.01$).

HYPOTHESES

1. Presentation Variables

It was hypothesized that the rate and level of learning a given skill varies as a function of the mode of model presentation.

a) Perceptual modeling procedures will produce significantly greater changes in the response strength of desired behaviors than will symbolic modeling procedures.

b) A combination of perceptual and symbolic modeling procedures will prove to be more effective than either procedure alone.

2. Feedback Variables

a) Prompting feedback will produce greater change than either self or confirmation feedback; the latter will produce the most.

b) A combination of prompting and confirmation feedback will prove to be more effective than any of the three forms alone.

3. Combinations of Presentation and Feedback Variables

a) The optimal combination of presentation and feedback procedures will be: a) perceptual and symbolic modeling in the presentation phase of training; and b) prompting and confirmation feedback in the feedback phase.

b) Initial gains in performance should be greater than in later phases of training. It was therefore predicted that performance gains following modeling treatments would be significantly greater than increases in Probing following feedback treatments.

THE DESIGN

A two by three design matrix that permitted an assessment of the relevant combinations of presentation and feedback variables yielded six experimental groups.

A. GENERAL PROCEDURE

The treatment procedure for each group was broken down into ten steps or stages (see Table 2). In six of these steps, all groups received identical treatment. In the remaining four, each group was exposed to the appropriate type of modeling and a particular type of feedback.

In the first step, all subjects were videotaped while teaching a five-minute lesson to four junior high school students. This constituted the pretest from which baseline levels of probing were derived.[1]

In step two all subjects received written instructions describing the criterion behavior. Following this set-induction process, the groups were exposed to the appropriate types of modeling in the next two steps. In steps five and six, all subjects planned and then taught a second five-minute lesson. The cycle was then repeated except that the set

---

[1] Five minute lessons were employed throughout the study. The rationale for this "microteaching" format derives from two considerations. First, Margolius and Sheffield (1961) found that in their training film research, a four to five-minute film segment turned out to be the optimal Demonstration-Attention segment for college-age military subjects. Further, McDonald, Allen and Orme (1966) found that if a videotape playback ran for twenty minutes, E's comments tended to be perceived as repetitious.

Secondly, the five-minute lesson provided ample opportunity for the intern or perceptual model to demonstrate a satisfactorily high number of probes for measurement purposes.

TABLE 2

SUMMARY OF STEPS IN TREATMENT BY EXPERIMENTAL GROUP (a)

	Phase and Step in Treatment	Experimental Group						Minutes in Treatment
		1	2	3	4	5	6	
	1. Teach (Pretest)	X	X	X	X	X	X	05
Modeling Treatments	2. Set Induction	X	X	X	X	X	X	20
	3. View The Self	A	E	A	E	A	E	10
	4. View the Model	R	R	A	A	E	E	10
	5. Plan Next Lesson	X	X	X	X	X	X	10
Feedback Treatments	6. Teach	X	X	X	X	X	X	05
	7. View The Self	A	E	A	E	A	E	10
	8. View The Model	R	R	A	A	E	E	10
	9. Plan Next Lesson	X	X	X	X	X	X	10
	10. Teach (Posttest)	X	X	X	X	X	X	05

(a) Explanation of Symbols: (X) indicates that the subject receives this treatment. (a) indicates that the subject views a playback of his own performance alone, or else views the perceptual model alone; (E) indicates that the subject views the tape concerned with the experimenter and thus receives discrimination training and reinforcement from him. (R) indicates that the subject restudies the written materials describing the criterion behavior.

induction process in step two was dropped. Note that in viewing the appropriate model for the second time, the subject was receiving a feedback treatment. The second presentation of the perceptual model constituted prompting feedback and symbolic modeling was defined as confirmation feedback.

In terms of general procedure then, all subjects were pretested, received written instructions, were exposed to the appropriate modeling procedure, and following a planning session, they taught for the second time. This means that before his treatment was complete, each subject had been exposed to the appropriate modeling and feedback treatments, and had taught three times.

To avoid undue attention being paid to the development of a new lesson each time, the subjects taught a different group of students in each lesson. This allowed them to retain the same basic subject matter in each lesson while attending to improvements in probing techniques.

The experiment was run over a six-week period as part of the regular intern program. Each subject spent approximately one hour and a half in total treatment.

B. TREATMENTS

As mentioned earlier, six experimental groups received differential treatments. Four groups viewed perceptual models at some point in their treatments and two received symbolic modeling.

To facilitate subsequent discussion, the groups will be identified by the kind of modeling they initially received. The overall treatment for each group will be considered in terms of presentation variables. Following this, feedback procedures will be discussed.

1. Presentation Treatments

A summary of the following discussion appears in Table 2.

Group One (Minimal Symbolic Modeling): Following the pretest and set induction through written materials (Appendix G), Group One subjects viewed a videotape playback of their pretest performance alone. Following this they were directed to study the written materials on the criterion behavior for a second time. In subsequent steps they planned and taught again, then viewed a playback of their second lesson, re-read the instructions once more and taught the third lesson.

Note that this is a symbolic modeling group. While they received written instructions,(no verbal instructions), discrimination training was provided by E. This group was thus termed a Minimal-Symbolic Modeling treatment group.

Group Two (Maximal Symbolic Modeling): Following the first two steps, subjects in this group viewed a videotape playback of their pretest performance with E. As in all treatment steps where he viewed a tape with the subject, E verbally reinforced the desired responses when they occurred during the playback. In addition he identified salient cues to which the desired behavior should be attached, made suggestions about variations in the form of the desired behavior, and pointed out the effects of such behavior on pupil behavior. In short, E's function was to provide appropriate reinforcement and discrimination training. Following the playback with E. Group Two subjects were directed to study the written materials for a second time. Then after a planning and teaching session, the cycle was repeated.

Note that like Group One, Group Two subjects received symbolic modeling. However, in addition to written instructions in step four, they also received verbal instruction from E in the form of discrimination training and reinforcement. This group was therefore termed a Maximal Symbolic Modeling treatment group.

Group Three (Minimal Perceptual Modeling): Following the first two steps, subjects in the third group viewed a videotape playback of their pretest performance alone. In the next step they viewed a perceptual model alone. As in all treatment steps where subjects viewed a perceptual model, Group Three interns were presented with a model teacher who was of the same sex as the subject, and who demonstrated Probing in a subject matter area which corresponded to the intern's major field of interest. For example, female English teachers viewed a female model teacher who demonstrated Probing in a five-minute English lesson.

Once Group Three subjects had viewed the model alone they planned and then taught the second lesson. The cycle was then repeated. Since Group Three constituted a perceptual modeling group who viewed the model without receiving discrimination training or reinforcement from E, they were defined as a Minimal Perceptual Modeling treatment group.

Group Four (Strong Symbolic, Minimal Perceptual Modeling): Like Group Two, this group received both discrimination training and reinforcement from E while viewing videotape playbacks of their own performance. This constituted a strong symbolic modeling procedure (Maximal Symbolic Modeling procedures required subjects to restudy the written instructions in addition to reinforcement and discrimination training).In the next step, subjects viewed the appropriate perceptual model alone. This treatment thus combined Strong Symbolic with Minimal Perceptual Modeling.

Group Five (Maximal Perceptual Modeling): This treatment differed from all other in that subjects viewed playbacks of their own performance alone, but viewed the appropriate perceptual model with E. His function was to provide discrimination training just as when he viewed a subject's playback with him. In this treatment however, his verbal output was keyed on the perceptual model's behavior rather than to the subject's performance.

This mode of treatment was termed Maximal Perceptual Modeling because in addition to viewing the model, subjects received discrimination training from E which was based directly on salient modeling cues, and served to increase the distinctiveness of these cues.

Group Six (Strong Symbolic and Maximal Perceptual Modeling): Subjects in the appropriate treatment stages viewed playbacks of their own performance with E, and then viewed the appropriate perceptual model with E. The condition was thus one of Strong Symbolic and Maximal Perceptual Modeling.

2. Feedback Treatments

By definition, all subjects received feedback treatments when they were exposed to the appropriate form of modeling for the second time. The second presentation of the perceptual model was defined as Prompting Feedback, and Symbolic Modeling became a form of Confirmation Feedback. A subject who viewed both his own playback and the perceptual model alone received Self-Feedback. The experimenter's function, when he viewed a tape with the subject, was precisely the same as in the earlier phases of treatment.

The relationship of each of these forms of feedback to the six experimental groups is outlined in Table 3. Groups One and Two did not view both the model and their own playbacks, and thus received only partial feedback. For this reason they were not considered as one of the four basic groups in the analysis of feedback effects.

C. SUBJECTS

Intern subjects were drawn from the Stanford Intern Teacher population. Prior to random assignment to the six groups, subjects were categorized by subject-matter major. From these subgroups, the interns were then assigned to one of the six treatment conditions. Foreign language majors were not included in the study, as their tapes would have presented undue measurement problems. They also follow a highly prescribed methodology that does not permit the systematic use of Probing techniques.

Relevant characteristics of the sample studied appear in Table 4. As can be seen, the groups are adequately matched on age, sex, ability and subject-matter variables. T tests for differences between groups were applied for each of the above variables, and proved to be non-significant.

Interns and models taught each of their lessons to a "micro-class" of four pupils. These students were drawn from the Palo Alto school system, and were paid for their services.

Forty students were hired to work in teams of four. Each team consisted of two boys and two girls. All pupils were either ninth or tenth

TABLE 3

SUMMARY OF MODELING AND FEEDBACK PHASES OF TREATMENT BY EXPERIMENTAL GROUP

Experimental Group	Phase of Treatment	
	Modeling (Steps 3 and 4)	Feedback (Steps 7 and 8)
1	Minimal Symbolic Modeling SM (min.)	----------
2	Maximal Symbolic Modeling SM (max.)	----------
3	Minimal Perceptual Modeling PM (min.)	Self-Feedback
4	Strong Symbolic-Minimal Perceptual Modeling SM (Str.) - PM (min.)	Confirmation Feedback
5	Maximal Perceptual Modeling PM (max.)	Prompting Feedback
6	Strong Symbolic-Maximal Perceptual Modeling SM (Str.) - PM (max.)	Confirmation and Prompting Feedback

TABLE 4

DESCRIPTIVE DATA FOR EXPERIMENTAL INTERN SUBJECTS
N = 111

Group Number	N	Mean Age	Verbal (GRE)	Quantitative (GRE)	Sex		Subject Matter Major						
					M	F	Eng	Soc. St	Math	Sc.	P.E.	Art	Drama
1	19	24.5	612.6	568.7	7	12	8	7	1	3	0	0	0
2	19	24.4	627.4	587.9	7	12	5	7	0	5	1	0	1
3	18	23.9	612.9	548.4	4	14	5	7	1	3	0	1	1
4	19	24.0	614.1	556.1	7	12	5	8	1	4	1	0	0
5	19	23.2	592.6	561.0	9	10	6	7	1	3	2	0	0
6	17	24.1	613.7	547.1	6	11	4	8	1	3	0	1	0

graders, and between 14 and 15 years old. The majority were from middle class background. However, approximately twenty percent were of lower-middle or upper-lower class extraction. None of the pupils received special instructions on how to behave in the microteaching situation. They were told only that Intern teachers would be teaching them various lessons. No mention was made of Probing.

D. TRAINING PROCEDURES FOR MODELS

Experienced teachers were selected to act as perceptual models. One model of each sex from each of the major subject matter areas was trained to demonstrate Probing techniques in a five-minute lesson. The conditions under which both models and interns taught were identical.

One week prior to taping and training, potential models were given an outline of Probing procedures (see Appendix G). In the training session they taught the same lesson to different groups of students until criterion was reached. The investigator set a lower limit of ten Probes for an appropriate lesson, and in addition required that the lesson be of "superior" educational quality in all other respects.

The latter decision was based on two sets of judgments. At the conclusion of each demonstration tape, the students filled out the Stanford Teacher Competence Appraisal Guide, an instrument that provides student evaluation across thirteen teacher behaviors on an eight-point scale (see Appendix K). In addition, the investigator and one other experienced teacher subjectively assessed the general educational quality of the tape. The models typically taught the same lesson three times. During each demonstration, the investigator recorded the number of Probes that occurred, and noted suggestions for improvement. In the period between each tape, the investigator and the model replanned the lesson. In all, eleven models received training. From this pool, the investigator selected the best seven tapes.

Models' characteristics are summarized in Table 5. It should be noted that the same female model was shown to female science and mathematics subjects. The lesson dealt with simple probability notions in science and was thus applicable to mathematics. Only three female mathematics subjects received perceptual modeling treatments, and since the intended female model in mathematics was weak, it was decided to forego further training with a new model and to use the science tape.

Two females in drama, and two in art viewed the female social studies model. Four males in physical education viewed the male English model. Since there were few subjects in these disciplines, models were not trained to demonstrate probing in these areas. The English and social studies models were selected in these cases because of all the tapes, their performances were the strongest in terms of number of Probes demonstrated.

In sum, eleven subjects viewed perceptual models who demonstrated Probing in subject matter areas other than the Intern's major area. However, the sex of the model was controlled in all cases.

E. TRAINING PROCEDURES FOR EXPERIMENTERS

Two experimenters were used in the study. As pointed out in an earlier section, they provided reinforcement and discrimination training at appropriate times in the treatments of the various groups. Both of the experimenters were male graduate students in Educational Psychology. Both had had prior experience with conditioning procedures and discrimination training in earlier research.

TABLE 5

SALIENT CHARACTERISTICS OF THE PERCEPTUAL MODELS AND THEIR RESPONSE FREQUENCIES FOR THE MAJOR RESPONSE CATEGORIES RATED

Model No.	Sex of Model	Subject Taught	Total Pupil Responses	Non-Probes	Probes	Clarify	Critical Awareness	Redirect	Prompt	Refocus	Model Repeats	Appraisal Guide Mean Scores (a)
1	Male	Soc. St.	18	4	14	5	1	7	1	0	6	58.2
2	Male	English	33	11	22	14	3	0	4	1	19	59.2
3	Male	Math.	16	6	10	6	2	0	2	0	8	60.0
4	Male	Science	25	6	19	10	3	2	4	0	6	61.2
5	Female	Soc. St.	20	5	15	8	2	4	0	1	10	63.0
6	Female	English	24	7	17	11	2	3	1	0	20	62.8
7	Female	Math/Sc.	26	10	16	4	4	1	3	4	19	57.0
Mean Frequency for all models			23.14	17.00	16.14	8.29	2.43	2.43	2.14	.86	12.57	
Proportions in Relation to Total Pupil Responses			1.00	.30	.70							
Proportions in Relation to Probes					1.00	.52	.15	.15	.13	.05		

(a) Pupils rated the model on an eight-point scale in terms of the following areas of competence: Aims; Planning; Performance; Evaluation. The lowest mean score a teacher could receive on this scale would be zero. The highest possible mean score would be 77.

Training was accomplished in the following manner. The author, who served as E_1, and E_2 jointly studied The Instruction to Interns (Appendix G), The Model's Manual (Appendix I) and The Rater's Manual (Appendix H). Following this E_1 and E_2 viewed the model tapes. In addition, E_2 observed E_1 modeling relevant discrimination training and reinforcement techniques with the first six experimental subjects. Finally, E_1 observed E_2 working with his first four subjects. As the experiment progressed, E_1 and E_2 periodically observed each other to ensure that they were employing equivalent procedures.

As each subject completed treatment, he was asked to complete a questionnaire which tapped his perceptions of the experimenter's ability, sensitivity to individual differences and general "likeability" (Appendix E) These results provide comparative data on subjects' perceptions of the experimenters, and are discussed in the Results section of this report. Experimenter effects are also discussed in that section.

In an attempt to minimize bias, neither experimenter was informed of the treatment conditions of the subjects with whom he worked.

F. METHODS OF GATHERING THE DATA

During the study, each of the interns' lessons was recorded on videotape for analysis. The relevant behaviors were later recorded by four raters trained for this purpose.

Prior to the analysis of the tapes in the current experiment, the raters had been trained and had rated approximately 400 twenty-minute tapes for another experiment in which the investigator was involved.

The dependent variable in this earlier study involved basic questioning techniques that included Probing among other things. Thus, the retraining phase for the current analysis was relatively brief. The raters retrained on nonexperimental tapes first, then rated the model tapes until ninety percent agreement was reached. Retraining included intensive study of The Rater's Manual and Coding Form (Appendix H), and practice on high frequency Probing lessons. Total retraining time was approximately six hours.

Once criterion was reached, the raters independently coded the experimental tapes. To maintain reliability, all four raters independently rated three experimental tapes out of each block of 50. Analysis sessions followed in which degree of agreement was discussed and reruns where necessary could be carried out. Note that these tapes were not used in the acquisition of reliability data; they served a maintenance function only.

G. INTER-RATER RELIABILITY

Interrater reliabilities on all types of responses rated proved to be very high. For this reason, and because further light is cast upon the nature of the dependent variable, the ensuing discussion will be fairly detailed.

The reliability data are based on independent, double ratings of all of the videotapes used in the study (N=326). The original plan, as in the Allen, McDonald, Orme (1966)[1] experiment employing Probing, was to randomly select a representative sample of tapes for double-rating.

---

[1] For ease in exposition, the Allen, McDonald, Orme (1966) experiment will hereafter be termed "Experiment II".

However, by the time that the raters had completed coding all of the tapes in the earlier study, they had become quick enough to rate a tape through with no holdups or reruns. Since the tapes in this study were only five minutes long, and since there were too few TV machines for each rater to work on a different tape, it was possible to have the tapes double-rated without incurring additional expense. Given two playback machines,four raters and one operator,it was found that by having the raters move from machine to machine (located in different rooms),one operator was engaged full-time playing tapes,changing them, and cleaning the machines. Operators played the tapes in a pre-established order, following a list in which groups and sessions were randomized (randomization based on Tables of Random Numbers). In this way, neither the operator nor the rater had any knowledge of the treatment condition or phase in treatment of subjects' tapes. Finally,to avoid the possibility that a disproportionate number of tapes,albeit independently rated,might be coded by the same two raters,a simple schedule of room changes was followed.

When the data were transferred to IBM cards, ratings on each tape were arbitrarily assigned to one of two data decks. Thus there were two complete sets of data on all subjects in all groups, the only differences between them being that they had been independently rated by different raters. Reliability coefficients were determined by running one complete deck against the other. Coefficients were determined for each of the major response categories and sub-categories rated. These data are reported in Table 6, along with the reliabilities obtained for the same variables in Experiment II (N=79).

Clearly, the coefficients are higher than those one has come to

TABLE 6

INTER-RATER RELIABILITY FOR EXPERIMENT TWO AND THE DISSERTATION DATA ON THE MAJOR RESPONSE CATEGORIES RATED[1]

Response Category	Experiment II (N = 79)	Dissertation Data (N = 326)
Total Pupil Responses	.99	.99
1. Pupil Questions	.73	.65
2. Pupil Comments	.87	.88
3. Pupil Answers	.96	.97
Intern Reinforces Pupil Response	.96	.99
Intern Repeats Pupil Response	.89	.98
Total Non-Probes	.99	.99
1. Intern Answers Question	.62	.80
2. Intern No Response	.73	.91
Total Intern Probes	.99	.99
1. Clarify	.97	.97
2. Critical Awareness	.87	.94
3. Redirect	.76	.97
4. Prompt	.55	.92
5. Refocus	.57	.87
*6. Encourages Alternatives	.59	---
*7. Summarizes	---	---
8. Role Play	---	---

*Response frequencies for these latter variables were too low to permit meaningful coefficients.

[1] Inter-rater agreement among all four raters on 20 tapes coded at different times throughout the sutdy was also determined. It was found to be above 90% on all response categories except for Clarification (89%) and Refocus (85%)

expect on the basis of reading the general literature on rating. It should be equally obvious that these results are due to extensive pre-training for raters, and unambiguous operational definitions of the variables involved(c.f., Appendix H: Rater's Manual). They are not a function of spuriously high agreement that may operate when raters work in teams, since they worked independently. It is also highly unlikely that pooling effects inflated reliabilities because ratings were not combined.[1]

It is possible that the Total Pupil Responses, Total Probing, and Total Non-Probing coefficients are somewhat inflated because they are made up of the summation of relevant sub-categories: i e.,

TOTAL PUPIL RESPONSES + TOTAL PROBES + TOTAL NON-PROBES

where, Total Pupil Responses = (Pupil Question) + (Pupil Rhetorical) + (Pupil Comments) + (Pupil Answers).

Total Probes = (Clarification) + (Critical Awareness) + (Redirect) + (Prompting) + (Refocus) + (Encourages Alternatives) + (Summarizes) + (Role Play)

Non-Probes = (Intern Answers Pupil Question) + (Intern No-Response) + (Intern Reinforces) + (Intern Repeats Pupil Response), provided that, in each case, the intern then fails to go on and Probe the pupil response--the latter qualification being necessary so that it would later be possible to determine whether a given sub-category in this class was employed as a Non-Probe, or used in conjunction with Probing.

---

[1] It should be pointed out here that all of the analyses reported in the next chapter were run on one data-deck only. This decision was made before reliability data were available. When measures of central tendency, distribution and variability were later run on both decks as a check, as might be expected from the high reliabilities, differences were miniscule. The "second" deck was therefore used only in the determination of reliability.

The proportionate contribution of each of the subcategories to its respective total, and the intercorrelations between them are relevant in trying to determine whether or not the summated or major response categories reflect spuriously high reliabilities. These data are presented in Table 7.[1]

In comparing the reliability coefficient for a given variable (Table 6) with its proportionate contribution to the relevant summated score (say, Total Pupil Responses in Table 7), it will be seen that there is a marked tendency for those responses which occurred infrequently to be less reliably rated. For example, Pupil Questions account for less than five percent of Total Pupil Responses, and at the same time proved to be the least reliably rated of such responses.[2]

In short, in view of the high reliabilities of the discrete response subcategories contributing most heavily to the summated categories of Probing, Non-Probing and Total Pupil Responses, it is unlikely that the latter are statistically unstable. One may well argue instead that they could be expected to be more stable than the lesser response categories and therefore to be preferred in statistical analysis.

---

[1] Certain other data are reported in Table 7 which are not relevant to the current discussion -- they will be discussed at a later time.

[2] Pupil Rhetorical Responses while included in the rating form as a possibly meaningful type of response, occurred so infrequently that analyses on this variable have no meaning. The same holds true for the Probing subcategories: Refocus; Encourages Alternatives; Summarizes; and Role Play.

TABLE 7

INTERCORRELATIONS AND PROPORTIONS FOR THE RESPONSE FREQUENCIES OF SUB-CATEGORIES OF PROBING AND PUPIL RESPONSES WITH TOTAL PROBES AND TOTAL PUPIL RESPONSES

Response Category	Perceptual Models (N = 7)	Expt. 2 Session 1 (N = 63)	All Session 1 (N = 109)	Experimental Session 2 (N = 109)	Subjects Session 3 (N = 109)	Intercorrelation Experiment 2 (N + 256)	Disser-tation (N = 326)
		Proportion In Relation to Total Pupil Response				r with Total P. Resp.	
Pupil Questions	----	.05	.04	.02	.01	.07	.12
Pupil Comments	----	.15	.20	.18	.19	.44	.53
Pupil Rhetorical	----	.01	.00	.01	.00	.03	.06
Pupil Answers	.98	.79	.76	.79	.80	.86	.89
Non-Probes	.30	.50	.53	.40	.38	.61	.63
Probes	.70	.50	.47	.60	.62	.79	.85
		Proportion in Relation to Total Probes (a)				r with Total Probes	
Clarification	.52	.64	.62	.57	.54	.85	.85
Critical Awareness	.15	.10	.10	.14	.13	.57	.50
Redirect	.15	.13	.12	.13	.18	.45	.42
Prompt	.13	.10	.14	.14	.13	.45	.56
Refocus	.05	.07	.04	.05	.05	.35	.34

(a)Summing all of the proportions of the sub-categories of Probing leads to a figure in excess of 100. This is due to the fact that it is possible to employ the Redirect techniques as a Clarifying context. When this occurred, a single response was rated as Redirect and Clarify. Such occurrences were infrequent, for example, overtotals are as follows: Experiment II. Session 1: 104; Experiment III Session 1: 102; Session 2: 103; Session 3: 103.

RESULTS

A. RATIO AND PROPORTIONATE ANALYSIS

Ratio analyses were performed on three different sets of scores. Probable modeling effects were explored by analyzing Session Two minus Session One mean scores ($S_2 - S_1$) and ($S_3 - S_1$) scores as well. In both cases Pretest or Session One scores were used as covariates. In addition, probable feedback effects were analyzed by comparing Session Three minus Session Two ($S_3 - S_2$) means. In the latter case both Session One and Session Two scores were used as covariates, i.e., pretest and modeling effects were employed as covariates so that feedback effects could be analyzed apart from these prior influences on performance. The results for all three analyses are presented in summary form in Tables 8 and 9. Significance was achieved on the ($S_2 - S_1$) analysis ($p < .05$), but fell short of the .05 level on the ($S_3 - S_1$)analysis ($p < .25$) where one would most likely expect significance. So also, the ($S_3 - S_2$) analysis yielded an F ratio that was non-significant ($p < .25$).

In the ($S_2 - S_1$) analysis, differences between Group 4 (SM - strl, PM - min) and the lowest scoring group (Group 1-SM min.) accounted for the significant F ratio. These results are inconsistent with all other results to be reported in two respects. First, differences where they occurred were most pronounced for ($S_3 - S_1$) analyses, not ($S_2 - S_1$) analyses as is the case for the Probe/Non-Probe analysis here. Secondly, the magnitude of mean differences in all other analyses led to group rankings which are generally consistent with the hypotheses (Group 6 highest, down through Group 1, the lowest). In the P/NP ratio analysis the rank order was instead, 4-2-6-5-1-3.

TABLE 8

SESSION TO SESSION TREATMENT MEAN (TRAINING) DIFFERENCES AND SIGNIFICANCE LEVELS WITHIN EACH GROUP ON THE MAJOR RESPONSE CATEGORIES RATED (a)

Variable	Direction of Differences	Experimental Groups 1	2	3	4	5	6
Total Pupil Responses	$(S_3>S_1)$	.01	.05	.01	.01	.01	.01
	$(S_2>S_1)$	NS	.01	.01	.01	.01	.01
Pupil Comments	$(S_3>S_1)$	NS	NS	NS	.10	NS	NS
	$(S_2>S_1)$	NS	NS	.10	NS	NS	NS
Pupil Answers	$(S_3>S_1)$	.05	.01	.01	.01	.01	.01
	$(S_2>S_1)$	NS	.01	.01	.01	.01	.01
Non-Probes	$(S_3>S_1)$	NS	NS	.05	.01	NS	NS
	$(S_2>S_1)$	NS	NS	.05	NS	NS	NS
Total Probes	$(S_3>S_1)$	.05	.01	.01	.01	.01	.01
	$(S_2>S_1)$	NS	.01	.01	.01	.01	.01
1. Clarify	$(S_3>S_1)$	NS	NS	.01	.10	.01	.01
	$(S_2>S_1)$	NS	.10	.05	.05	.05	.01
2. Critical Awareness	$(S_3>S_1)$	NS	NS	.05	NS	NS	.01
	$(S_2>S_1)$	NS	.05	NS	.05	.05	.05
3. Redirect	$(S_3>S_1)$	NS	.05	.05	.10	NS	NS
	$(S_2>S_1)$	NS	NS	.05	.05	.10	NS
4. Prompt	$(S_3>S_1)$	NS	.05	.05	.05	NS	NS
	$(S_2>S_1)$	NS	NS	.01	NS	NS	NS
5. Refocus	$(S_3>S_1)$	NS	NS	.05	NS	NS	.05
	$(S_2>S_1)$	NS	NS	NS	NS	NS	.05

(a) The studentized range or T statistic was used in the comparison of treatment mean differences (Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, pp. 89-92.

TABLE 9

SUMMARY OF THE ANALYSES OF VARIANCE FOR ADJUSTED MEAN DIFFERENCES BETWEEN ALL GROUPS OVER ALL SESSIONS WITH PROBES OVER NON-PROBES (P/NP) AS THE DEPENDENT VARIABLE

Type of Analysis		Experimental Group						N	df	F-Ratio[(a)]
		1	2	3	4	5	6			
$(S_3 - S_1)$	Adjusted Mean	0.33	0.68	1.03	0.65	1.50	1.16	108	5,101	1.390[a]
	Standard Error	0.37	0.35	0.37	-.35	0.35	0.37			
$(S_2 - S_1)$	Adjusted Mean	0.55	0.81	0.52	2.51	0.68	0.70	107	5,100	2.387**
	Standard Error	0.49	0.49	0.51	0.49	0.48	0.53			
$(S_3 - S_2)$	Adjusted Mean	-0.59	-0.29	0.07	-0.60	0.57	0.21	105	5,97	1.669[a]
	Standard Error	0.37	0.35	0.38	0.37	0.35	0.38			

(a) Level of Significance: **: p .05
a: $p < .25$

Similar analyses of Total Probes/Total Pupil Responses were also run. Results here were consistent with other analyses, but fell short of significance. However, the $(S_3 - S_1)$: $F_{5,101} = 1.797$;($p < .25$), and $(S_3 - S_2)$: $F_{5.97} = 1.653$; ($p < .25$) analyses showed a trend in the predicted direction.

In view of the questionable nature of the distributions of ratios (P/NP) and proportions (P/TPR), these results are suspect. Perhaps the most meaningful approach to the problem of ceiling effects lies in a straightforward though inelegant inspection of the shifts in treatment means for each group as it moves through the three experimental sessions. These shifts are depicted in Figure 1.

It is clear that the proportion of Probes to Total Pupil Responses systematically increases for all groups from Session One to Session Two. This is also the case from Session Two to Session Three, except that Group 1 (SM - min.) and Group 4 (SM - str., PM - min.) show a decrease. The dropoff for Group 4 is particularly dramatic. In contradistinction to the ratio analysis, the data here would suggest that if significant differences were going to occur, they would show up most clearly in an $(S_3 - S_1)$ rather than an $(S_2 - S_3)$ analysis.

B. MODELING EFFECTS $(S_3 - S_1)$ and $(S_2 - S_3)$ ANALYSES OF COVARIANCE.

Table 10 presents a summary of the analyses of covariance for between group differences in all sessions with Total Probes as the dependent variable. While the F ratio for differences showed a trend in the predicted direction by Session Two ($p < .25$), it will be seen that a second training or treatment session was necessary before the differential effectiveness of certain treatments became sufficiently

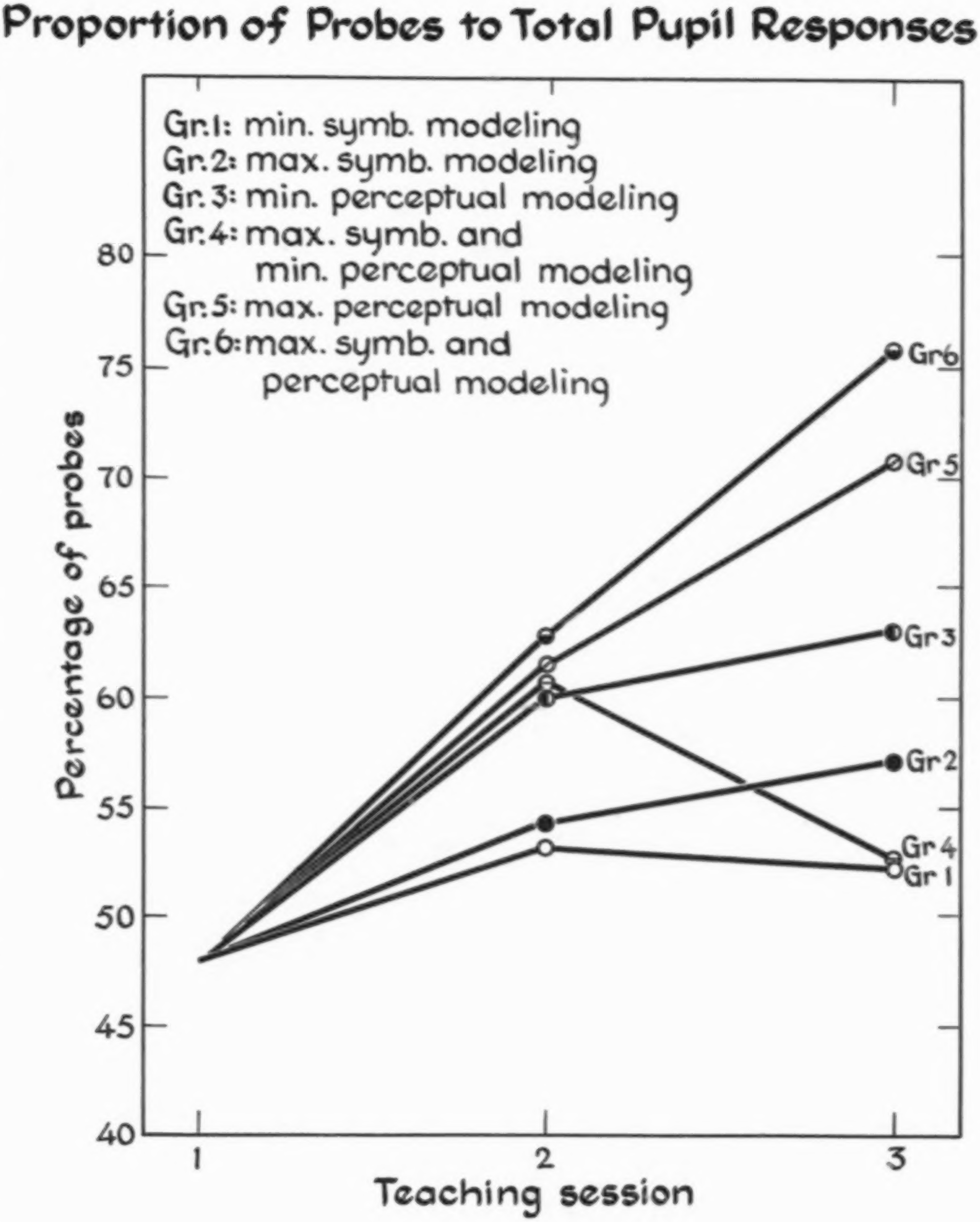

Figure 1, Experiment III

Adjusted Mean Frequency of Proportion of Probes to Total Pupil Responses Across Teaching Sessions for Each Experimental Group.

TABLE 10

SUMMARY OF THE ANALYSES OF VARIANCE FOR ADJUSTED MEAN DIFFERENCES BETWEEN ALL GROUPS OVER ALL SESSIONS WITH TOTAL PROBES AS THE DEPENDENT VARIABLE

Type of Analysis		Experimental Group						N	df	F-Ratio
		1	2	3	4	5	6			
$(S_3 - S_1)$	Adjusted Mean	4.71	5.81	8.01	6.39	8.04	9.72	108	5,101	2.571^{**}
	Standard Error	1.14	1.08	1.15	1.08	1.08	1.13			
$(S_2 - S_1)$	Adjusted Mean	3.64	5.44	5.98	6.58	7.31	8.01	107	5,100	1.544^{a}
	Standard Error	1.22	1.19	1.27	1.22	1.19	1.29			
$(S_3 - S_2)$	Adjusted Mean	-0.10	0.01	1.49	0.12	1.27	2.64	105	5,97	1.233
	Standard Error	0.98	0.90	0.99	0.92	0.90	0.99			

(a) Level of Significance: **: $p < .05$
a: $p \leq .25$

powerful to be reflected at an acceptable level of significance (F - 2.571; $p < .05$).

These modeling effects are illustrated in Figure 2. The pretest performance of each group on Probing was used in the adjustment for initial differences. The plot-points for each of the groups in Session Two are derived from the adjusted means differences between Session Two and Session One frequencies of Probes; those for Session Three on the (S_3 - S_1) adjusted mean differences. Since each group started from a pretest score which obviously could not differ from itself, the curve for each group begins at zero, and thus the means entered in the figure did not have to be brought to a common point by the application of a constant. Needless to say, this is a real advantage in trying to present a meaningful picture of the results.

Note that from Session One to Session Two, the groups perform as expected. However, by Session Three, Group 4 (SM - str., Pm - min.) and Group 3 (MS - min.) have changed their relative positions on the predicted treatment continuum. Both groups, it will be remembered, were exposed to perceptual models. They differed from each other in that Group 4 received discrimination training from E when viewing the self, whereas Group 3 viewed both model and self-playbacks alone.

Comparable analyses with Clarification as the dependent variable were also performed, and yielded results that correspond closely with those for the Probing analysis. A summary of the results of the covariance analyses carried out on Clarification appear in Table 11, and are illustrated in Figure 3.

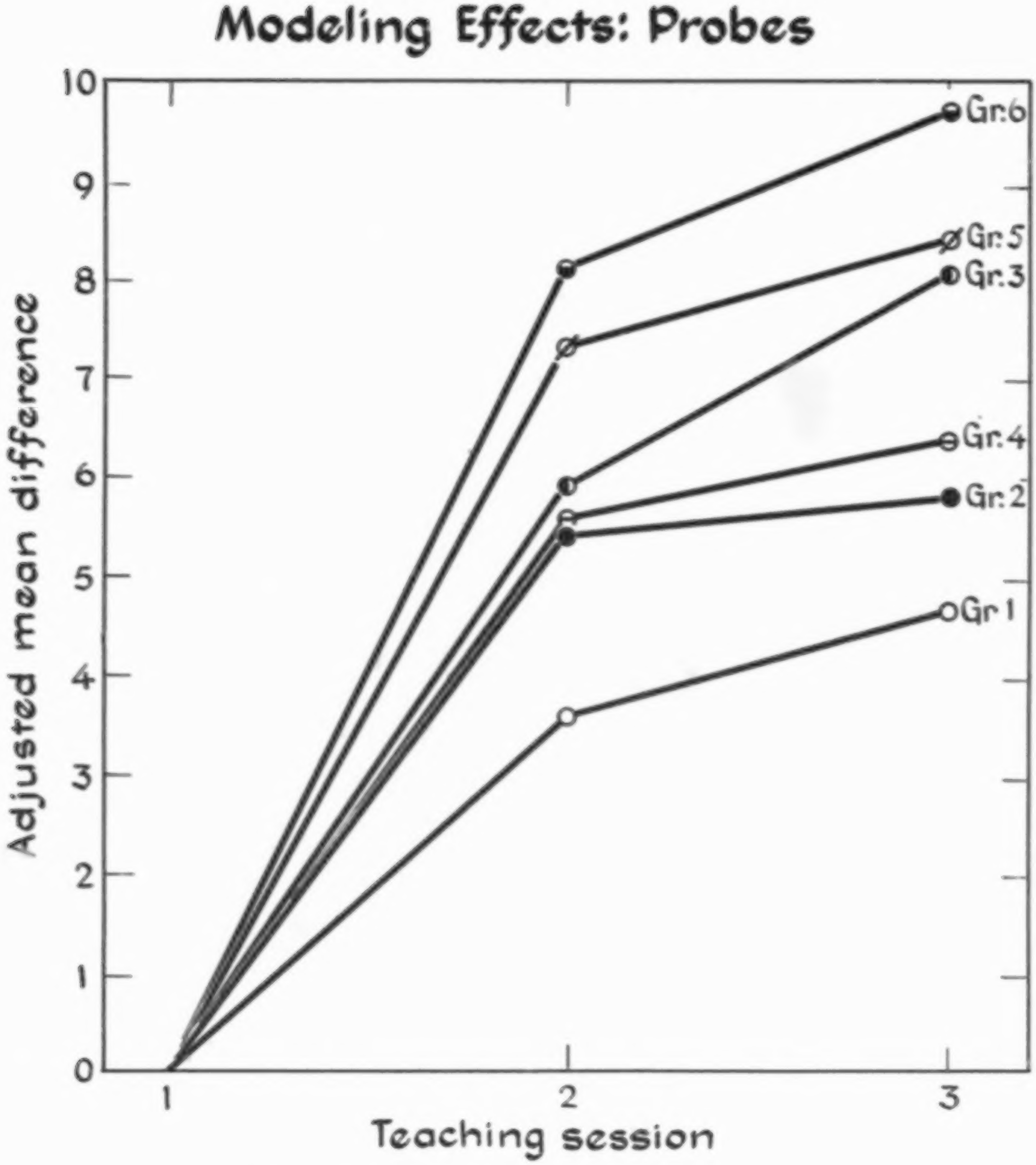

Figure 2, Experiment III

Adjusted Mean Differences Between Teaching Sessions
2 and 1, and Sessions 3 and 1, for Each Experimental Group.

TABLE 11

SUMMARY OF THE ANALYSES OF VARIANCE FOR ADJUSTED MEAN DIFFERENCES BETWEEN ALL GROUPS OVER ALL SESSIONS WITH CLARIFY AS THE DEPENDENT VARIABLE

Type of Analysis		Experimental Group								
		1	2	3	4	5	6	N	df	F-Ratio$^{(a)}$
$(S_3 - S_1)$	Adjusted Mean	2.21	2.30	5.19	3.20	4.59	5.85	108	5,101	.962***
	Standard Error	0.09	0.86	0.91	0.86	0.86	0.91			
$(S_2 - S_1)$	Adjusted Mean	2.21	2.92	4.07	3.57	3.84	4.35	107	5,100	0.705
	Standard Error	0.93	0.91	0.96	0.93	0.91	0.99			
$(S_3 - S_2)$	Adjusted Mean	-0.55	-0.86	1.01	-0.30	0.92	2.17	105	5,97	2.457**
	Standard Error	0.74	0.70	0.76	0.71	0.69	0.76			

(a) Level of Significance: ***: $p < .025$
**: $p < .05$

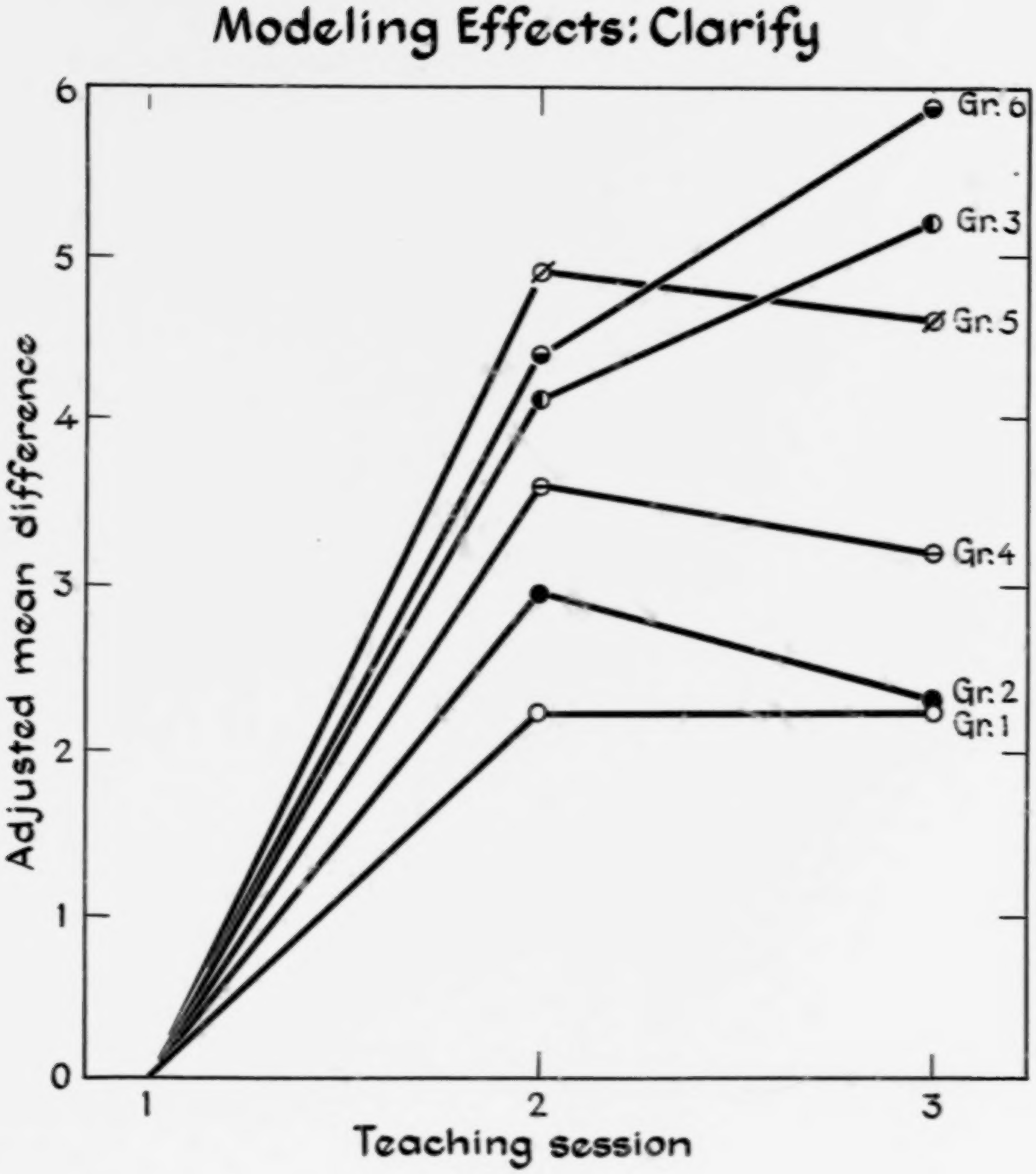

Figure 3, Experiment III

Adjusted Mean Differences Between Sessions 2 and 1, and Sessions 3 and 1, for Each Experimental Group.

The most discernable differences between the two sets of analyses involve Groups 2 (Sm - max.) and 4 (SM - str., PM - min.) both of whom show decreases in response strengths on Clarification which are not reflected in Probing responses. Table L-1[1] presents data which appear to resolve this issue. It will be noted that both of these groups show proportionate increases in the sub-categories of Probing which counter-balance shifts in Clarification. For Group 4, there is almost a 10 percent gain in the lesser sub-categories from Session One to Two while Clarification in relation to Total Probes drops correspondingly. This is enough to change the rank order of this group. This shift, however, is not a significant one.

Similarly, Group 2 subjects demonstrated increases in Probing sub-categories from Session Two to Three, such that they show a fall off in Clarification, but retain a slight increase in Total Probes.

Before specific tests on the hypotheses are reported, other general treatment differences must be discussed. Table 12 presents a summary of F ratios and significance levels for analyses conducted on all of the major variables involved in the study. It is evident that the only significant increases in response strength by Session Two were those for Pupil Answers ($p < .05$). These are closely related to Total Pupil Responses ($r_{PA \cdot TPR} = .89$), and together with a trend ($p < 25$) in Pupil Comments were nearly enough to produce significant gains in the broader category (p for Total Pupil Responses $< .10$).

A summary of the analyses for Pupil Answers appears in Table 13. Figure 4, which plots these adjusted mean differences, is very similar to

[1] See Appendix L.

TABLE 12

SUMMARY OF F RATIOS AND SIGNIFICANCE LEVELS FROM THE ANALYSES OF COVARIANCE PERFORMED ON EACH OF THE MAJOR RESPONSE CATEGORIES FOR SESSIONS TWO AND THREE

Dependent Variable	Adjusted Mean Differences					
	$S_3 - S_1$ (N = 108)		$S_2 - S_1$ (N = 107)		$S_3 - S_2$ (N = 105)	
	$F_{(5,101)}$	Significance	$F_{(5,100)}$	Significance	$F_{(5,97)}$	Significance
Pupil Comments	0.416	N.S.	1.469	.25		
Pupil Answers	2.171	.10	2.341	.05		
Total Pupil Responses	1.815	.25	2.138	.10	0.587	N.S.
Intern Repeats Pupil Responses	2.492	.05	2.084	.10	1.573	.25
Total Non-Probes	0.633	N.S.	1.558	.25	0.667	N.S.
Total Probes	2.571	.05	1.544	.25	1.223	N.S.
Clarification	2.962	.025	0.705	N.S.	2.457	.05
Critical Awareness	0.637	N.S.	1.243	N.S.	0.787	N.S.
Redirection	0.376	N.S.	1.861	.25	0.312	N.S.
Prompting	0.574	N.S.	1.142	N.S.	0.748	N.S.
Refocus	0.678	N.S.	2.276	.10	0.438	N.S.

TABLE 13

SUMMARY OF THE ANALYSES OF VARIANCE FOR ADJUSTED MEAN DIFFERENCES BETWEEN ALL GROUPS OVER ALL SESSIONS WITH PUPIL ANSWERS AS THE DEPENDENT VARIABLE

Type of Analysis		Experimental Group						N	df	F-Ratio[a]
		1	2	3	4	5	6			
$(S_3 - S_1)$	Adjusted Mean	6.23	6.17	8.80	7.75	9.28	10.92	108	5,101	2.171*
	Standard Error	1.29	1.21	1.31	1.21	1.21	1.28			
$(S_2 - S_1)$	Adjusted Mean	3.99	5.86	6.11	6.49	8.29	9.92	107	5,100	2.341**
	Standard Error	1.33	1.28	1.39	1.32	1.28	1.40			

[a] Level of Significance: **: $p < 05$
*: $p < .10$

that for Probes (Figure 2). This is not surprising in view of the strength of the association between these two variables ($r_{P.P_A} = .91$).

An inspection of the F ratios for Total Probes and its sub-categories along with Non-Probes in the $(S_2 - S_1)$ analysis implies considerable response variability in the second session. The rank-order of adjusted mean differences for Total Probes ($p<.25$) is exactly as predicted. Howevery, for those sub-categories which approach significance or show trends in the right direction, the order of treatments is somewhat variable, and it will be noted that there is also a discernable increase in Non-Probes ($p<.25$). By Session 3, however, rank-orderings appear to have been stabilized.

The response category termed Intern Repeats Pupil Responses (Repeats) surprisingly approaches significance in the $(S_2 - S_1)$ analysis ($p<.10$) and in the $(S_3 - S_1)$ analysis, exceeds the .05 level of significance. A summary of the analyses of covariance and the relationship of adjusted mean differences for each group with all others appears in Table 14. These means behave in a manner analagous to Probing variables.

This was surprising because the experimenters' sets about this variable were that it would occur fairly frequently during the pre-session tests, and if not extinguished or suppressed during training, would tend to "crowd-out" Probing responses. They were quite wrong on both counts. Intern Repeats were found to be more highly correlated with Probing (.65) than with Non-Probes (.35), and response strength increased significantly from Session One to Session Three, rather than decreasing.

For the Experiment Two data mentioned earlier, in which subjects received treatments like those for Group Two in this study but taught

TABLE 14

SUMMARY OF THE ANALYSES OF VARIANCE FOR ADJUSTED MEAN DIFFERENCES BETWEEN ALL GROUPS OVER ALL SESSIONS WITH INTERN REPEATS AS THE DEPENDENT VARIABLE

Type of Analysis		Experimental Group						N	df	F-Ratio (a)
		1	2	3	4	5	6			
$(S_3 - S_1)$	Adjusted Mean	3.38	2.69	5.27	4.55	4.37	7.07	108	5,101	2.492^{**}
	Standard Error	0.99	0.94	0.99	0.94	0.94	0.99			
$(S_2 - S_1)$	Adjusted Mean	2.05	2.02	4.23	2.92	4.32	5.96	107	5,100	2.084^{*}
	Standard Error	1.04	1.02	1.09	1.05	1.02	1.11			
$(S_3 - S_2)$	Adjusted Mean	0.54	-0.18	1.18	1.43	0.51	3.00	105	5,97	1.573^{a}
	Standard Error	0.86	0.82	0.89	0.83	0.81	0.89			

(a)Level of Significance:
**: $p < .05$
*: $p < .10$
a: $p < .25$

regular lessons in their classroom, Intern Repeats correlated .54 with Total Probes, and .26 with Non-Probes. In addition, while initial response strength on Repeats appeared somewhat higher to begin with, there were no significant changes during training. For Experiment III on the other hand, all of those groups who were exposed to perceptual models during treatment showed significant gains both in the $(S2 - S_1)$ and $(S_3 - S_1)$ contrasts for training differences with the T statistic (see Table 8). The Symbolic Modeling Groups, like those in Experiment Two showed no such gains.

There seems to be little doubt then that these differences are due to modeling treatments, and are not explainable by other variables. Certainly experimenter influence is not operating in favor of Groups 3, 4, 5, and 6. If anything it could be expected to attenuate the results. An inspection of the model protocals in Table 5 shows that Repeats occurred in considerable strength. The possibility is that Intern Repeats became an S_D which served to trigger Probing responses, i.e., repeating the pupil response gave the intern or model a brief time period during which he could relate a given probe to the subject-matter at hand.

More simply, and perhaps in combination with the above, through contiguous association with Probing as exemplified in the models' performance, Repeats became conditioned to certain pupil responses. If such were the case, then through modeling procedures, stimuli which had formerly been bound up with other verbal habit patterns came to act as cues to signal other behaviors such as Probing.

C. FEEDBACK EFFECTS

Overall treatment differences for feedback effects were analyzed by considering (S_3 - S_2) mean differences with two covariates. As in the analyses for modeling effects, each of the major response categories were considered, but the focus now shifted to an examination of predicted differences between Self-Feedback (Group 3), Confirmation (Group 4), Prompting (Group 5), and Combined Prompting and Confirmation (Group 6) procedures.

An inspection of Table 12 data shows that significant differences were achieved only on the Clarification analyses. While Intern Repeats showed a trend ($p < .25$), all other analyses yielded non-significant F ratios.

The summary of the analyses of covariance for feedback effects (S_3 - S_2) on Clarification is presented in Table 11. Figure 4 presents adjusted mean differences for the analysis on this variable and includes adjusted mean differences (S_3 - S_2) on Total Probes as well. The latter means are included to provide an illustrative comparison between the two response categories when modeling effects are controlled for by covariance adjustements. It will be seen that predicted order effects for these latter analyses are wide of the mark in that Group 3 (Self-Feedback) demonstrated greater gains than either Prompting or Confirmation Feedback Groups. This is somewhat surprising in view of results obtained in Experiment I in which teacher reinforcement of pupil participation was the dependent variable , and the Self-Feedback group showed decrements in response strength for Sessions Three and Four. The response pattern here is precisely the opposite.

D. INTERVENING VARIABLES ANALYSIS

As was pointed out earlier, T statistic comparisons among means for potential experimenter, age, and GRE performance differences were tested and proved to be non-significant. A more sophisticated analysis in which analyses of covariance on Session Three means for selected major response categories (Session One mean scores as covariates) were also run. In these analyses for general linear hypotheses on all possible differences between say, Experimenter 1 and Experimenter 2, on a given dependent variable such as Probing or Non-Probing, F ratios for row, column and interaction (row x column) effects are reported. This allows one to determine not only whether there were experimenter (row) effects on a given response category for any one of the six experimental groups, but also to obtain a second covariance analysis on F ratios with different df on main effects, i.e., even though row F ratios may be non-significant, thus telling us that there were no differential experimenter effects, column F ratios will reflect the significance of difference between means on say, Total Probes across the six groups. These ratios will tend to be smaller, since the df are usually considerably less. For example, in the analysis for potential experimenter effects, adjusted S_3 means for Groups 2, 4, 5, and 6 (those groups who were exposed to E at some point in treatment) were analyzed where Total Probes, Non-Probes and Intern Reinforces the Pupil adjusted means were entered into the analysis. The df for row or experimenter effects were thus 1 and 62, those for column (treatment) effects 3 and 62, while those for row x column effects where one might find that a given experimenter was differentially effective with say, Group Six subjects, but not Group Two subjects, were also 3 and 62. Here, then, in column effects is a supplementary analysis for treatment differences with a reduced N and fewer degrees of freedom.

Before these results are summarized, the rationale for selecting

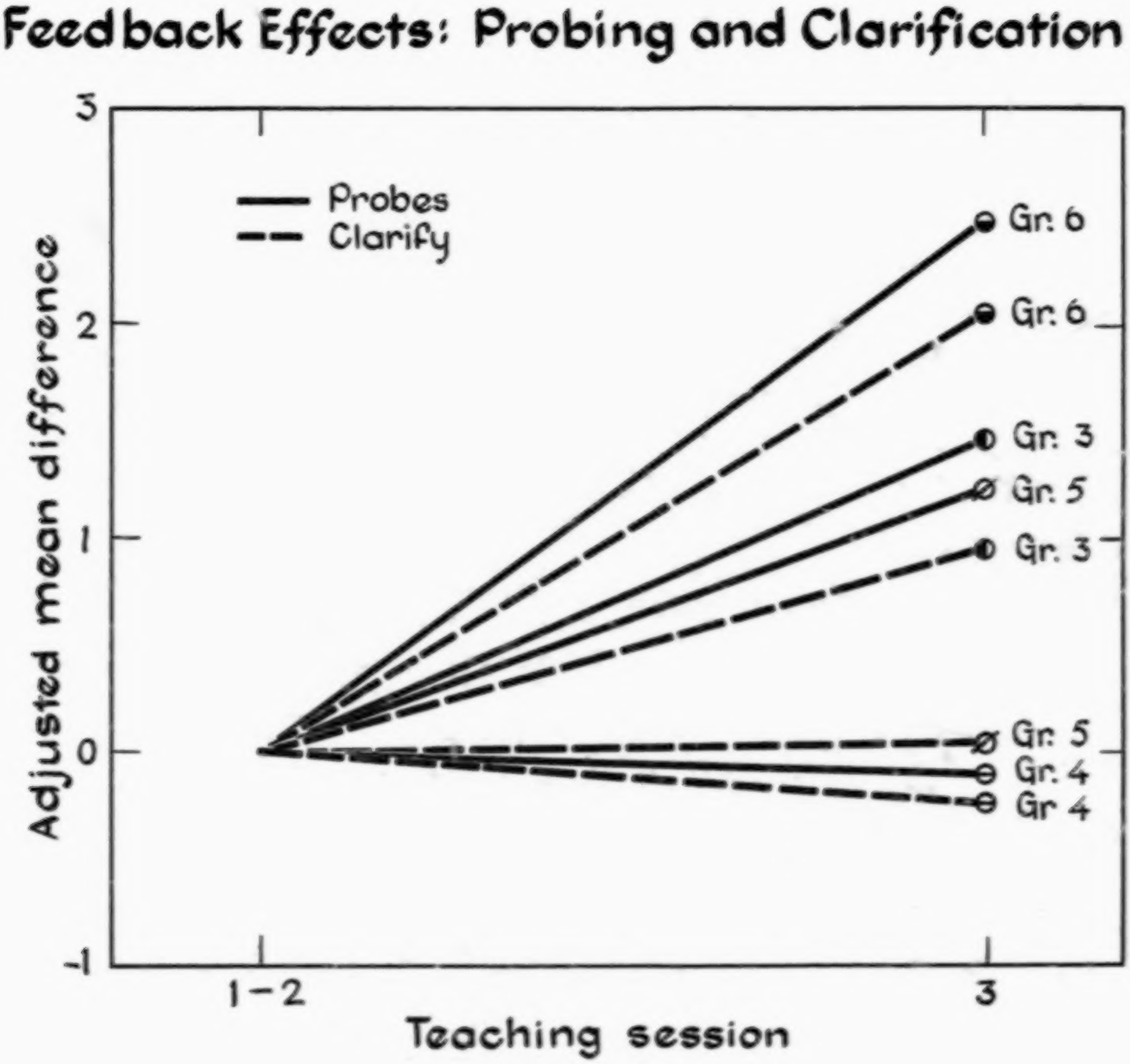

Figure 4, Experiment III

Adjusted Mean Differences Between Sessions 3 and 2 for Probe and Clarification Responses for Each Experimental Group.

Reinforcement, Probes, and Non-Probes for the analysis of intervening variable effects will be briefly outlined.

Probes and Non-Probes were obvious choices because they constitute general measures of the dependent variable. An examination of Table 12 will also show that Probes produced significant differences in the $(S_3 - S_1)$ analysis, and yielded trend F ratios $(p < .25)$ in the other two types of analysis. Thus they seemed a reasonable choice. Non-Probes, of course, are the logical opposites of Probes, and for this reason were also included. Reinforces was selected as the third variable because E_1 had had considerable experimental experience in this area, and if differences apart from those on the major response categories were likely, Reinforces might well reflect them. In addition, reinforcement procedures are known to be responsive to sex differences, and since they are predominantly verbal behaviors, it was thought that they should also pick up GRE verbal score differences if any existed.

Of all of the analyses for potentially differential effects: by experimenter, sex, perceptual model, GRE verbal and GRE non-verbal scores, only one significant result was obtained. Subjects in each group were dichotomized at the median into a high or low scoring group in terms of their GRE Verbal scores. The analyses of covariance on the three selected dependent variables showed that by the third session, low scorers did not Probe significantly more frequently than did the high scoring group $(F_{1,94}) = 5.239$; $p < .05$). The low scorers were not, however, localized in any particular experimental group, for interaction analysis yielded a non-significant F ratio. The inference that the low scoring group may also have Probed significantly less frequently than the high scorers is incorrect, since the row F ratio on Probes was non-significant. Column effects for Probing were, however, significant ($F_{5,94}=2.699$; $p<.05$), providing further support for the(S_3 - S_1) analysis on Probing reported earlier.

There were no significant sex differences. However, here again, column effects for Probing were significant ($F_{5,95} = 2.735$; $p<.05$). Similarly, analyses for the differential effectiveness of one or more models yielded nonsignificant differences. Column effects for Probing were also non-significant ($F_{3,59} = 1.549$; $p<.25$).

Finally, GRE quantitative scores and experimenter effects proved to be non-significant, although added support for significant differences between the experimental groups on Probing was again found ($F_{3,62}=2.834$; $p<.05$) on the experimenter-effects analysis.

Having presented evidence which indicates that the significant differences between groups on the covariance analyses for major response categories of dependent variables are not due to non-treatment effects such as differential experimenter performance, intelligence or other variables of a similar nature, this discussion now turns to specific tests of the hypotheses.

E. SPECIFIC TESTS OF THE HYPOTHESES

The selection of the most appropriate test for specific contrasts following significant F ratios is problematic, and requires brief discussion.

There is general consensus among statisticians that the most appropriate test for *a priori* hypothese following significant F ratios is the *t* test (c.f. McNemar, 1962; Winer, 1962; Ferguson, 1959). However, it is also known that when *t* is used for multiple comparisons among means, the probability of committing a Type-1 error is approximately .05. This suggests that, in these cases a more rigorous basis than usual should be required for the rejection of the null hypothesis. One could avoid the *t* test altogether and use the Tukey Test or the Scheffé method. But as Winer (1962, p. 89) points out, these *a posteriori* tests may lead to the commission of Type-2 errors, even when one is "data-snooping." An additional complication involving the T test is that it requires equal N's and this would mean an appreciable reduction in sample size for the Experiment III data as N varies from 17 to 20, i.e., 54 cases would have had to be discarded in such an analysis.

Following Ferguson (1959, p. 238) the rejection region for the *t* statistic was examined in terms of the more rigorous 10/K (K-1) percent level, where K is equal to the number of groups. In terms of Experiment III data, the adjustment was such that in order for the *t* to be reported as being significant at the .05 level, it would actually have had to reach the .017 level in conventional tables for critical values of *t*.

Rather than working out the adjustment for all other appropriate significance levels, it was decided to double the required level of significance for a given *t* value. Thus in Table 15, where the significance level for *t* is reported to be at the .05 level, it is actually at the .01 level in terms of conventional tables for *t*. Since the appropriate *t* test for *a priori* hypotheses is for one-tail of the distribution only, the 10/K (K-1) adjustment has been made by reading off all relevant values for two-tail tests, and reporting them as being one-tailed tests. This procedure may thus be seen to be slightly biased in favor of rejecting

TABLE 15

BETWEEN-GROUP (TREATMENT) COMPARISONS AND CONTRASTS FOR DIFFERENCES BETWEEN ADJUSTED MEANS FOLLOWING SIGNIFICANT F RATIOS BY ANALYSES OF COVARIANCE

Response Category and Type of Analysis	Direction of Differences		Student's *t* for df of 100	Adjusted Significance Level[a]
Probing ($S_3 - S_1$) :	Group 6	Group 1	3.158	.005
	Group 6	Group 2	2.495	.01
F = 2.57	Group 6	Group 4	2.148	.025
5, 101	Group 5	Group 1	2.159	.025
	Group 5	Group 2	1.501	.10
	Group 3	Group 1	2.111	.025
	Group 3	Group 2	1.400	.10
Clarification	Group 6	Group 1	2.881	.005
($S_3 - S_1$):	Group 6	Group 2	2.816	.005
	Group 6	Group 4	2.319	.025
F = 2.962	Group 5	Group 1	1.938	.05
5, 101	Group 5	Group 2	1.853	.05
	Group 3	Group 2	2.319	.025
Intern Repeats Pupil	Group 6	Group 2	3.221	.005
Response ($S_3 - S_1$):	Group 6	Group 1	2.678	.005
F 5,101 = 2.492	Group 6	Group 5	2.013	.025
	Group 5	Group 2	1.337	.10
	Group 3	Group 2	1.900	.05
Pupil Answers	Group 6	Group 1	3.134	.005
($S_2 - S_1$):	Group 6	Group 2	2.178	.025
F 5,100 = 2.341	Group 5	Group 1	2.378	.025
	Group 5	Group 1	1.378	.10
Clarification	Group 6	Group 2	3.015	.005
($S_3 - S_2$):	Group 6	Group 1	2.682	.005
	Group 6	Group 4	2.494	.01
	Group 3	Group 2	1.180	.15

[a] The adjusted probability levels reported above are based on a more rigorous rejection level (10/k (k-1), where K is the number of groups) than that for conventional *p* values (c.f. Ferguson, 1959, pp 237-238). The adjusted p values are for one-tailed tests.

a given hypothesis when in fact it should be accepted.

It does, however, avoid the extremes of the T test on one hand and the unadjusted t test on the other.

Hypothesis 1 (a): states that, "Perceptual Modeling procedures will produce significantly greater changes in the response strength of desired behaviors (Probing responses) than will Symbolic Modeling procedures." The statistical tests of this hypothesis are therefore t tests in which Group Three (Minimal Perceptual Modeling) is pitted against Group One (Minimal Symbolic Modeling), and Group Five (Maximal Perceptual Modeling), is compared with Group Two (Maximal Symbolic Modeling).

Table 15 which reports the results for these and all other contrasts show that t test results for $(S_3 - S_1)$ mean differences among the relevant groups on Total Probes, generally support the hypothesis.

Group Three (PM min.) subjects, as predicted showed significantly greater gains than did Group One (SM min.) subjects (t for df of 100 = 2.111; $p < .025$, one-tailed). Group Three subjects also tended to be significantly different from the Maximal Symbolic Modeling subjects (Group Two) on Total Probes ($t = 1.40$; $p < .10$; one-tailed). Maximal Perceptual Modeling procedures (Group Five) led to significantly greater performance gains on Total Probes than did the Minimal Symbolic Modeling treatment ($p < .025$), and tended to be superior to Maximal Symbolic Modeling (Group Two) as well ($p < .10$).

The contrasts on adjusted means for Clarification as the dependent variable led to similar results, though at greater significance levels. Again, Group Three showed significantly greater gains in the desired behavior than did Group One ($p < .01$), and was also significantly superior to Group Two ($p < .025$). Group Five (PM min.) was again significantly different from Group One (SM min.) ($p < .05$) and unlike the analysis for Probes where it fell just short of significance, in Clarification, Group Five proved to be significantly different from Group Two ($p < .05$).

Added support for the above findings may be seen in the results for Intern Repeats. As was explained earlier, differences were not predicted for this variable, but are due to treatment effects, and move in the same direction (Group Five - Group Two, $p < .10$; Group Three - Group Two, $p < .05$). Finally, as Figure 1 shows, when ceiling limits on Probing are taken into account, the differences are consistently in favor of the hypothesis even though probability levels may not be safely stated as the means are unadjusted.

The results then are quite consistent. With the effects of prior experience partialled out, it was found that by the Third Session, Minimal and Maximal Perceptual Modeling treatments were generally more effective than equivalent Symbolic Modeling procedures in increasing Clarification responses in particular and Probing in general.

The results are attenuated somewhat by the lack of significance between the Maximal Symbolic and Perceptual Modeling conditions on Total Probes. It is concluded then that the first hypothesis is partially supported by the data.

Hypothesis 1 (b): states that a combination of Perceptual and Symbolic Modeling procedures will prove to be more effective than either Perceptual or Symbolic procedures alone.

The test for this hypothesis required that comparisons be made between Group Six (SM - Str., PM - Max.) and Groups Five, Three, Two and One. Groups Five and Three represent Maximal and Minimal Perceptual Modeling respectively, and Groups Two and One are their counterparts in Symbolic Modeling.

Table 15 summarizes the t test results for these comparisons. It was found that on the $(S_3 - S_1)$ analysis, Group Six consistently and significantly showed greater gains on Total Probes, Clarification, and Intern Repeats than did either of the Symbolic Modeling conditions. These differences as shown, exceed the .01 probability level.

While Group Six was also significantly different from Group Five on Intern Repeats ($p < .025$), the former did not in general differ from either of the Perceptual Modeling Groups in question.

It is concluded that the second hypothesis (1b) is partially supported by the data. Combined Perceptual and Symbolic Modeling procedures proved to be significantly more effective than did Minimal or Maximal Symbolic Modeling treatments alone. There is no consistent evidence to indicate that the Combined procedure is superior to either Minimal or Maximal Perceptual Modeling alone.

Hypotheses 2 (a) and 2 (b): relate to feedback procedures, and state that Combined Prompting and Confirmation Feedback (Group Six) will yield greater gains in the desired behavior than either Prompting Feedback (Group Five), Confirmation Feedback (Group Four), or Self-Feedback (Group Three).

The only analysis relevant to these considerations is the $(S_3 - S_2)$ analysis in which feedback effects are considered independently of the main analysis for modeling effects by covariance adjustments for prior effects (S_1) and initial modeling effects (S_2). Of the relevant analyses of covariance that were carried out in this manner, only that on Clarification proved to be significant. The *t* test for comparisons among means supported only one of the sub-hypotheses: Combined Prompting and Confirmation Feedback (Group Six) proved to be significantly different ($p < .01$) from the Confirmation procedure (Group Four).

Group Three (Self-Feedback) subjects performed well beyond expectation. In fact, the adjusted means for the groups on this analysis show that predictions were generally wide of the mark.

The hypotheses on feedback effects are therefore not supported by the data.

Hypothesis 3 (a) is a general hypothesis with particular relevance for training experiments where the experimenter wishes to identify the optimally effective treatment condition. It states that the optimal combination of presentation and feedback procedures will be one in which Combined Symbolic and Perceptual Modeling procedures in the presentation phase are combined with the Prompting and Confirmation Feedback condition (Group Six).

A complete evaluation of the hypothesis requires an examination of (S_3 - S_1) differences on all relevant variables entering into Probing. Table 15 shows that Group Six subjects consistently and significantly realized greater gains on Total Probes, Clarification, and Intern Repeats, than did Groups One, Two or Four. Group Six also showed significant gains over Group Five on Intern Repeats. These differences extend from the .005 to .025 level.

The differences between Group Six and Groups Three and Five are not significant on the main variables. However, as Figures 1 through 4 amply demonstrate, Group Six subjects uniformly showed the greatest gains, and achieved the highest mean scores.

Hypothesis 3 (b) is a general hypothesis relating to performance curves, and states that performance gains from Session One to Session Two will be significantly greater than those from Session Two to Session Three, regardless of experimental group.

To test this hypothesis, adjusted mean difference scores for all groups from Session One to Session Two on Probing were averaged and then compared with the overall mean difference of the groups for Session Two to Session Three. The t test for correlated means was a highly significant 21.32 ($p < .005$, two-tailed). An inspection of the data indicated that a similar analysis of scores on the Clarification variable would have yielded even larger differences, so no further tests were carried out.

An examination of the figures already presented will show that as predicted, the gains from Session One to Session Two on all of the major variables (excluding Non-Probes) are demonstrably greater than those from Session Two to Session Three. It is therefore concluded that the hypothesis is supported by the data.

DISCUSSION OF RESULTS AND IMPLICATIONS FOR FURTHER RESEARCH

The general assumption from which hypotheses were derived was that the rate and level of learning a given teaching strategy varies as a function of the mode of model presentation. Specific predictions were based on theoretical considerations which suggest that the differential effectiveness of varying model and associated feedback procedures stems from their distinctive cueing properties.

The effects of Symbolic and Perceptual modeling together with certain feedback conditions were explored in terms of their ability to increase the distinctiveness of relevant cues. Training procedures were varied along a continuum of increasingly available cues and information on the criterion behavior.

Beginning with a common base of written instructions that described the desired behavior, the minimally effective condition consisted of repeated exposures to the written instructions, and pictorial representations of subjects' prior performance. The power of successive treatments was increased by the addition of a verbal discrimination training on relevant cues and the inclusion of perceptual models who portrayed the desired responses.

The treatment continuum extended then from a condition in which the criterion behavior was described through written instructions only (Minimal Symbolic Modeling), to the addition of verbal instructions on relevant cues (Maximal Symbolic Modeling), and then to procedures where the desired responses were actually portrayed by filmed teacher-models (Minimal Perceptual Modeling). Finally, the clarity of cues and the amount of information were manipulated by combining the filmed demonstrations with additional cue-discrimination training in the form of verbal instructions from E (Maximal Perceptual Modeling).

Since verbal instructions from E included both modeling (verbal reinstatement of model characteristics) and feedback (reinforcement of correct responses) functions, the cueing continuum was further extended to include treatments where these verbal instructions were linked with either the Perceptual Model's performance, or S's prior performance, or both. The former highlighted E's modeling *function;* verbal instructions based on S's prior performance emphasized E's feedback foie, and the combined procedure represented the most powerful combination of those cueing procedures that were considered.

Between Group Differences: Following initial written instructions common to all groups, the Perceptual Modeling treatments led to significantly greater gains in Probing techniques than did the Symbolic Modeling conditions.

Perhaps the most striking differences among all groups were those between Maximal Symbolic Modeling (Group Two), and Minimal Perceptual Modeling (Group Three). Following set induction, Group Two S's received discrimination training based on playbacks of their prior performance. In addition, they restudied the written instructions immediately following the individualized instruction from E on these behaviors. Group Three S's on the other hand received no discrimination training.

Following set induction, they simply viewed playbacks of their prior performance, then the model tape, alone.

In establishing a treatment continuum on theoretical grounds, these two groups fell on either side of the symbolic-perceptual demarcation line. The differences between them were significantly greater than for any other immediately adjacent pair of groups. Group Three S's achieved significantly greater gains than did Group Two S's on Clarification ($p < .025$), Intern Repeats ($p < .05$) and tended to maintain these differences on Total Probes ($p < .10$) in the ($S_3 - S_1$) analysis. These differences support the distinctions made between the two types of modeling on the basis of their cueing properties.

A notable difference between the Maximum Symbolic (Group Two) and Perceptual (Group Five) modeling conditions in terms of the ease of training should be mentioned. Considerably less finesse and effort were required of E when he provided model-based discrimination training than when his comments were based on subjects' tapes. In the Perceptual Modeling conditions, a high frequency of Probing behaviors occurred and were modeled in a clear-cut manner. This increased the distinctiveness of relevant cues and discrimination training under these conditions appears to have been more efficient.

Finally, in terms of treatment differences between specific groups, it was concluded that Group Six, which represents the most powerful combination of Symbolic and Perceptual Modeling procedures, constitutes the optimal training procedure. The evidence in support of this conclusion is drawn from the consistent superiority of this group across all types of analysis.

Certain qualifications do,however, attach to specific hypotheses, and these must now be considered. Experimentally,the fact that all six treatment conditions produced significant changes in the response strength of the dependent variable could well have turned out to be a mixed blessing. The major consideration in research of this kind is the differential effectiveness of various treatment conditions. If the least powerful procedure produces changes which bring subjects close to the asymptote of performance on a series of behaviors with a ceiling limit, then the experimental utility of the dependent variable is severely curtailed, as potentially significant differences among treatments are wiped out. While this did not happen in this experiment, such considerations are relevant, for on the treatment continuum from Minimal Symbolic to Combined Maximal Symbolic and Perceptual Modeling (Group One through Group Six), those groups immediately adjacent to each other and within each set of modeling treatments did not as a general rule differ significantly.

The empirical nature of the dependent variable appears to have been an important factor here. What appears to have happened is that the greatest gains in Probing techniques occurred in Clarification, which was at the same time, the most frequently occurring Probing behavior prior to training. The evidence indicates that in the thick of classroom interaction, the Probing techniques which are most likely to be used frequently, are those which are already prepotent in teacher's response repertoire. Unless training is directed specifically towards shifts in this response hierarchy, increases in Probing regardless of their scope will follow the original distribution of responses. Had the models received discrimination training more directly focused on each of the sub-categories of Probing, rather than upon Probing in general, it is likely that treatment differences would have been correspondingly sharper.

These outcomes underscore the experimental utility of Bandura's (1963a,c) designs, where models are trained to emit highly unusual behaviors. The contingencies in training research are, however, quite different, and it may frequently prove to be impractical to seek out unique operants to serve as dependent variables. Viable teaching skills and strategies can be expected to have a certain "survival value", having been tested out over the years, and in developing and employing them as dependent variables, one can expect them to be occurring in some strength prior to treatment. At the very least, their nature is such as to make them relatively "easy" to learn, since the central verbal elements of these skills will already be a part of the response-repertoires of adult subjects. The operant strengths of likely skills and strategies should be assessed with naive populations before experimentation begins.

Finally, the above re-examination of Probing as a basic questioning technique clearly implies that pre-lesson planning must be highly systematic and concrete if the Probing which occurs during interaction is going to go beyond the obvious. Probing is not a simple skill that can be easily acquired. It might be pointed out in this regard that even though approximately one-half of their responses could be classified in the Clarification category, learning curves based on group means by session did not approach asymptote by the second session as they apparently did in Experiment I where teacher reinforcement for selected pupil responses was the dependent variable. Indeed, indications are that in the current experiment, further trials would have led to further gains, as the learning curves for each group on Total Probes (Figure 2) were still on the rise in the Third Session.

General Implications for Further Research: This study represents an initial attempt to adapt modeling procedures to teacher training. It is obvious that the last word is yet to be said on this topic.

The question of the transferability of the skills learned under the training conditions outlined here is a central issue, and is yet to be systematically explored. There is, for example, evidence to indicate that the more closely one is able to approximate actual performance conditions in training, the more effectively will such procedures contribute to later performance (Roshal, 1961). This suggests that in replications of this kind of study, experimental groups be included in which interns or teachers receive modeling experiences while teaching regular classes, and that models be presented teaching in comparable conditions.

Variables such as the amount of practice, the optimal number of cues emitted by models, sex differences in models and the sequencing of steps in treatment were controlled in this study rather than systematically varied, and appear to be worthy of future investigation. In addition, the provision of multiple models in training constitutes an inviting research possibility. Interns in the present experiment stated that they found the second exposure to the same model somewhat repetitious. One could hypothesize that either: (a) the provision of two or more different models demonstrating the same techniques under standard conditions would enhance learning because they were better able to maintain attention to salient cues, and because they provided greater variety in the initial presentation of tasks; or, (b) one could hypothesize that the provision of multiple models would impede the rate of learning because of their initial stimulus-novelty value (leading the subject to cue-in on irrelevant details). Such associations must be allowed to extinguish before discrimination on the relevant cues can begin to increase in strength.

Finally, the nature of the dependent variable is such that it would probably be worthwhile to assess its relevance in terms of performance criteria. It seems to have considerable face validity. The question remains however, does the use of such techniques produce demonstrable changes in pupil performance?

Generalizability of the Data: There are at least two general questions one might ask in attempting to generalize the results of this study. First, what is the probability that the results obtained here would hold for other teacher samples under analagous experimental (or training) conditions.

Excluding the foreign-language teachers, the sample for this study constituted the entire Stanford Intern class. Their performance on the Graduate Record Examination marks them as a relatively select group. Their responses to the questionnaire, and frequent informal communications to E and other personnel connected with their training but not with the experiment, suggest that they were well-motivated, and perceived the training as highly useful. Most importantly perhaps, all of the subjects were beginning teachers, and had not been long in the program. They had not as yet been given even limited responsibility for a classroom of students.

It is possible that individual differences such as those between neophytes and experienced teachers could act in such a way as to produce differences among the Perceptual Modeling treatments. For example, Minimal Perceptual Modeling might prove to be less effective with experienced teachers than with neophytes because the former may have developed response patterns which are highly resistant to extinction. In such cases, treatment conditions such as those for Group Four where Minimal Perceptual Modeling was supplemented with discrimination training (or retraining) based on subject playbacks might prove to be more effective.

It is not, however, the intent of this closing discussion to generate new rationales. The above illustrations are simply included to show that first, the results of the present study must be viewed within the limitations set by the design and secondly, the study may serve a heuristic function in generating new areas for research.

The second issue that is relevant to considerations of generalizability, relates rather more directly to the kinds of skills or strategies one might wish to transmit to teachers. While Perceptual Modeling techniques proved to be superior to written or verbal instructions, where basic questioning techniques were involved, it is doubtful that this superiority would be maintained across all teacher behaviors. It is not unlikely that certain behaviors can most efficiently be transmitted simply through verbal or written instructions.

On the other hand, there is some evidence to indicate that in certain situations, real-life models may be more effective than filmed models. While Bandura, Ross and Ross (1963b) found that filmed and real-life models led to comparable changes in aggressive behavior, recent work by Bandura and his associates (1965) indicates that in cases where emotional responses of the subject are central to behavior change (e.g., fear of dogs; phobic behavior) real-life models produce greater response shifts than do filmed models.[1] Filmed portrayals would appear to be particularly effective where non-verbal behaviors involving gestural responses are relevant to the acquisition of a skill, or where the S^D that triggers the desired behavior is clearly discriminable as was the case in this study, and where the behavior is sufficiently complex - (say, conducting a chorus) that precise written and verbal instructions are inadequate.

---

[1] It should be pointed out here that in distinguishing between Real-Life and Symbolic Models, Bandura and Walters (1963b, p. 49) consider filmed demonstrations of a given behavior or strategy to be examplars of the latter.

By varying the distinctiveness of cues from written instructions to filmed demonstrations, and not considering Real-Life models in the subsequent classification of treatments, this distinction was blurred in the current study.

All of the treatments in this study constitute forms of Symbolic Modeling, so that the two central forms considered might well have been labeled Symbolic-Symbolic and Symbolic-Perceptual.
This latter procedure corresponds more closely with the current literature on modeling as a whole.

CHAPTER V

CONCLUSIONS

Conclusions: Finally, it is concluded that the results of this study support the assumption that the rate and level of learning a given teaching strategy vary as a function of the mode of model presentation. More specifically, there is evidence to indicate that Perceptual Modeling procedures are characterized by distinctive cuing properties which tend to recommend them over Symbolic Modeling procedures for use in training contexts analagous to those described in the experiment.

Due to the lack of clear-cut support for all of the experimental hypotheses, these conclusions are tempered with the qualification that while treatment differences occurred as predicted, they were in the main concentrated at the extremes between the two types of modeling.

The results of these experiments may be summarized in terms of(1) their contributions to the analysis of teaching behavior and its effects; (2) the information they yield about the effects of training variables on producing certain categories of teaching behavior; (3) the problems they suggest. Only the second of these categories is the direct yield of the experimentation. The first and third categories represent observations made during the course of experimentation and data analysis. Not all of the comments made in these two categories have been tested in the strict sense. They are offered here to alert our colleagues to methods and research questions which may interest them.

The Significance of Experimenting on Teaching Behavior.

Those experiments represent the few,if not the only,experiment on teaching behavior in which the conditions of a true experiment have been maintained. They represent a *proof in principle* that such experimentation can be conducted.

Thus, the possibility develops that we may now test theories of teaching in the same rigorous way that some other empirical theories have been tested. The vision of an empirical science of teaching promises to be realizable. The development of such a science through systematic experimenting will inevitably change the nature of teaching practice. Three moderately successful experiments is a small seed from which to grow that flower of hope and expectation. But the demonstration that such experiments can be done successfully was the first necessary step which, until it had been taken, made visions of the hopes of a rigorous science of teaching.

The Technology of Experimenting on Teaching Behavior.

The portable videotape recorder was the single most significant contributor in directing these experiments. This device provided a means of providing feedback on a trainee's behavior in a way that no other means provided.

Human observation and comment has been largely ineffective for many reasons, principally that the observer is a limited data-gatherer, and the interaction between trainee and observer is fraught with problems. Films and other types of video equipment are cumbersome and the time-lapse to viewing is greater (this latter characteristic may be less significant than we originally believed).

The portable video recorder makes instant playback a reality, is unobtrusive after familiarizing teachers and their students with it, and is flexible in use, easily moved (an important characteristic when trainees are teaching in widely separated locales), durable, and relatively economical.

None of these characteristics, however, would have had great psychological significance if the process of viewing oneself on videotape did not act as an effective stimulus. Although we cannot prove it, it seems that such viewing is analogous to seeing oneself as one performs. The trainee seems to recreate the teaching experience, with sufficient distance, however, to objectify it to some degree.

Also, the video recording is an objective record available both to the trainee and independent observers. Although a trainee may distort for himself what the record may seem to portray to another, this kind of a record cannot be dismissed or distorted as easily as the reports of observers on the actions of teachers.

These two aspects of the video recording, its power to recreate the original experience and its objectivity, appear to provide the necessary conditions for using the video record as a training device. How the record may be used is an important problem for study. In the experiments reported here the video record was used as part of a feedback system and as a way of portraying the desired teaching behavior.

Results of Experiments

The results of the experiments described in detail in the preceding sections may be summarized under the following headings: (1) what were the most effective training variables; (2) what seemed to be the least effective variables; (3) what results appear to be inconsistent with other psychological research.

An important disclaimer is necessary here. Significant training effects were obtained in all experiments; i.e., trainees improved on the criterion behavior across teaching sessions in all treatments. Even though there is an unaccounted for drop in some groups in the last session, all groups performed at a higher rate at the end of the experiment than at the beginning.

There is no "pure" control group in any of these experiments, a group that is tested but receives no relevant training. Our control groups all received some type of training which was relevant to, but not necessarily specific to, the behavior to be learned. For example, in Experiment I, a group viewed its own performances on videotapes and evaluated them on the dependent variable (reinforcing students' participating behavior). This group's treatment differed from that of the other groups in that trainees' attention was not focussed on the dependent variable. Similarly, in Experiment III, all groups received some form of modeling instructions, no group receiving no such demonstration. The "weakest" treatment consisted in a detailed description of the desired behavior.

In each of these cases it is apparent that all trainees in these experiments were receiving some form of treatment relevant to the behavior to be acquired. Since we do not at present know the specific factors which may produce the desired behavior changes, it is not unlikely that some aspects of a critical variable, or, perhaps, even the critical variable itself, may be included in these supposedly "weak" treatments. For example, viewing oneself on videotape may be the necessary condition for producing certain behavior changes. A person may need to see himself behaving to conceptualize how he is behaving and to contrast a desired performance with his observed performance.

Thus, in Experiment I, a trainee might have no idea of the manner in which and the frequency with which he reinforces students' participating in class. Seeing himself on videotape may be a necessary cue to alert him to the need for and ways of modifying his behavior. When the behavior being learned is a common social behavior, such as rewarding other people's verbal behavior, the simple viewing of one's own performances may be sufficient to initiate a behavior change. Although we obtained significant differences across treatments in Experiment I, the line of reasoning presented here suggests that these differences might have been even larger had we used a "no-treatment" control group.

On the other hand, when, as in Experiment III, the behavior to be learned may not require observing one's own performance, differences are even less likely to emerge. In that experiment, the behavior to be learned may be described by giving a rule: "Whenever a student offers an answer in class, ask him another question which requires him to go beyond the answer that he has just given." It seems plausible that a trainee could follow this rule rather easily without having observed his own performance to determine whether he is already following it.

Observing his own performance may cue him to instances when he failed to observe the rule. Observing a model's performance may cue him to the variety of forms applying the rule may take. Thus, self-viewing and viewing a model's performance enhances learning, but the critical learning occurs once the subject has grasped the rule.

The point of this line of argumentation is that the results of the presented research demonstrate that some variables are probably effective, but the reader is to be aware that the relative effectiveness of these variables is probably underestimated in some cases, and in others, may not have been. The summary of the results described in the following section should be evaluated in the light of these qualifications.

Most Effective Variables:

It seems clear that the single most effective variable was a form of self-viewing, accompanied by prompting by an experimenter during the self-viewing. The most powerful treatment in Experiment I was one in which the trainee viewed his own performance and received reinforcement and cue discrimination training from an experimenter. Similarly, the most effective treatments in Experiment III involved the use of an experimenter in a feedback condition.

The most effective variable for describing a desired behavior appears to be a modeling condition in which the behavior is portrayed, and in which the subject views the model's performance while being cued by an experimenter on the significant aspects of the model's behavior. The results of Experiment III support this statement. In that experiment, any condition in which the experimenter was present while the subject viewed a model tape was always more effective than those conditions in which the subject viewed the tape alone. The single most effective condition combined a modeling experience with a feedback condition, both of which had an experimenter present who cued the trainee on desired behavior.

Although the experiment has not been performed which effectively separates out the effects of these two kinds of variables, it seems clear that for producing some kinds of behavior change, a modeling and feedback condition with an experimenter present during both phases is a powerful treatment.

As pointed out previously, this result does not appear strongly enough in the experimental results to accept it without qualifications. Such a treatment might be demonstrated to be highly effective when the behavior being learned is not easily cued by simple instructions. It also seems likely that a feedback condition will probably be more effective in producing those behaviors for which the trainee already has component responses in his repertoire.

Least Effective Variables:

The results of Experiment II, although they are difficult to interpret and must not be interpreted as a conclusion (since no significant differences were obtained), are consistent with a fact that was apparent in the first two experiments. In Experiment II, time-lapses between the time of the occurrence of the original behavior and the time when the subject actually viewed his performance were uncorrelated with measured changes in behavior, even though these time lapses extended over days and weeks. Again, repeating the warning not to interpret the acceptance of the null hypothesis as demonstrating an experimental effect, these results are so striking that they must raise a question about an important variable--the immediacy of feedback.

In a teaching situation, the teacher receives a certain amount of feedback upon his performance from the behavior of students. This information is received more or less immediately after the enactment of the teacher's behavior. For example, if the teacher asks a question and receives an appropriate answer, the teacher assumes that his presentation to this point has been effective. Similarly, other types of student behavior, such as attention and interest, give the teacher information on his teaching behavior.

On the other hand, when supervisors evaluate a teacher's performance, such information is usually given at a point in time after the behavior has actually occurred. Ordinarily, the shortest time between the enactment of the behavior and feedback from a supervisor on that behavior is the time between when the behavior occurs and the end of a teaching period. In the chapter on theory for this report, we pointed out the limitations of giving feedback verbally about behavior which occurred at an earlier point in time. However, the use of the videotape recorder suffers from a similar limitation--it can be used only after the actual teaching itself has been completed.

However, unlike the verbal report, it appears that this time-lapse is not a critical factor. In both Experiment I and II trainees received information on their teaching performance hours, and in the case of trainees in Experiment II, days and weeks after the actual performance. Experiment I indicates that this time lapse was not a critical factor. In that experiment the time lapse was equated across these treatments. In all treatments, feedback was delayed by several hours, yet striking changes occurred. It might be argued that the results would have been greater had the feedback been closer in time. The plausibility of this argument is weakened by the results of Experiment II. In Experiment II, where the time-lapse was systematically varied, but the training was held constant, no effective differences were obtained. When the results of these two experiments are put together in this fashion, it appears that the immediacy of the feedback, where immediacy is measured in terms of time, is not a critical factor.

The explanation for this may be that the videotape playback reinstates the trainee's performance for him. The whole experience of viewing oneself on the videotape is quite different from receiving information from a second person about one's performance. The character of the feedback experience has changed drastically. Whatever factors may be involved in this new experience are sufficiently different so that the factor of immediacy is no longer relevant.

This observation, if confirmed in further experimentation, is both theoretically and practically important. Theoretically, it suggests that feedback processes depend essentially on the information processing characteristics of the subject, not on the time-space relations of the feedback to the subject. Practically, reducing or eliminating the need to provide immediate feedback makes a wide range of feedback systems possible, which are easier to manage and more economical.

Consistency with Results of Other Psychological Experimentation:

It is clear from the results of Experiment I, with the exception noted in the immediately preceding section, that the feedback variable is highly effective in changing behavior. As was pointed out earlier, reinforcing desired behavior produces behavior changes. The problem is to determine what kinds of reinforcements, in what amount, and dispensed at what rate, are most effective in changing which kinds of teaching behavior. Practically, a simple decision rule seems to be: *Always include a feedback system in which the trainee views his own performance with supervision*.

It is important to note that conditions in which the feedback was provided without the assistance of an experimenter were always less effective than those in which an experimenter was present. It is difficult in this study to separate out the reinforcing properties of the experimenter's behavior and its cue-discrimination characteristics. Thus, we have a problem which has faced psychologists for many years. Although we have preferred to think of the feedback system as essentially information-providing, it is apparent, from naturalistic observations during the experiment, that this information has reinforcing characteristics which are only moderately correlated and may be uncorrelated with the information characteristics of the feedback.

It may be that feedback is most effective when it occurs closest to the time the behavior is to be performed next. This possibility makes sense when we recall that a teaching behavior is performed at one time and may not be repeatable until the next day. This hypothesis is easily tested.

The modeling treatments in Experiment III were not as powerful as we had hoped. We suspect that the critical factor in this case is the nature of the dependent variable. Probing behavior is a common, if not frequent, form of verbal interaction, particularly among well-educated people, like the trainees in this experiment. Furthermore, these trainees have been exposed to many models of this same behavior in their own very recent college experience, and perhaps may be perceiving models of it during their current training period. It is not unusual for university teachers to query students on their thinking. The "give-and-take" of many a seminar consists in exchanges of probing-type questions. Since our experiment is a "real-life" experience in the sense that the behavior to be learned is to be translated immediately into "real-life" situations -- (actual teaching), and since it is observed and used in everyday experiences, it may be that the effectiveness of modeling has simply been attenuated. Modeling may be more effective with behavior changes where the behavior to be learned is less readily observable and infrequently practiced.

It should also be pointed out that our most effective modeling treatments involved an experimenter present to cue the subject on the desired behavior. We think that this fact is important for the theoretical development and understanding of observational learning. In earlier experimentation, the dependent variable, such as aggressive behavior, may have been so dramatic, or so sharply contrasted with other on-going social behavior, that the observer had no difficulty detecting this behavior. However, trainees observing a complex verbal interchange on a complex set of ideas between a teacher and his students, may not easily or readily detect the specific form of the teacher's behavior which is occurring consistently. For example, a probing question may take many different forms, even though its general characteristics and timing remain the same. With an experimenter present to point this out,the trainee quickly learns what behavior to look for and what salient characteristics to identify.

If this line of reasoning proves to be empirically sound, we will then move the study of observational learning to studies of the interaction between characteristics of observational learning and characteristics of the behavior to be learned. In the study of teaching behavior, this would mean that we would have to study what kinds of teacher behavior need to be cued when modeled and what kinds do not.

Suggestions for Future Research:

We are proceeding on a plan of research to study various conditions of modeling and reinforcement. It seems obvious that further progress can be made by studying which characteristics of models make the modeling procedure more effective, and which characteristics of the feedback procedure make it more effective. For example, it is possible to present both positive and negative models of a desired

It is also possible to present both positive and negative feedback on the trainee's performance as he views it. Both of these are instances of characteristics of the modeling and feedback treatment which need to be investigated.

The most significant line of research suggested by the results seems to be the study of the interactions between the kind of teaching behavior to be learned and the kind of treatment used to produce the behavior change. Although our studies do not demonstrate the fact, it seems apparent that some types of treatment are more likely to be effective with specific kinds of teaching behavior and not with others. The first approach to the analysis of this problem will be to contrast our most effective treatments across the series of dependent variables which we have developed.

Still another line of potentially fruitful investigation would be to relate the characteristics of the trainee to the effectiveness of the treatment. We observed what appeared to be differential reactions to a common treatment. It was obvious that some subjects liked some treatments and some did not. How these differences may be related to the characteristics of the trainees is presently unknown.

Studies need to be done on the behavior of the experimenter during an experimental treatment. We proceeded largely by applying common sense, using a standard form of verbal interaction to mediate our treatments. However, it may be that some experimenter behaviors interacted with the specifics of the treatment to render it more or less effective.

It is also obvious that the personality characteristics of the experimenter are themselves a set of variables which probably subtly influence the treatment. We do not know, for example, if male experimenters should be paired with male and female, or male alone, or female alone, trainees. (We also do not know whether or not the sex of a model has an effect on the modeling treatment).

Again on the basis of our naturalistic observations, we wonder if certain classes of teaching behavior as we have developed them can be, or will be, learned by both men and women trainees. The probing-style seems more 'natural' to a male teacher than to a female teacher. We may have been attempting to produce behavior changes which are less likely to be learned by women than by men. Although this is a problem to be studied, even if the results were to concur with the suggestion we have just made, the implication would not be that the female trainee should not acquire these teaching behaviors. The obvious problem for research is to determine how to change behaviors when attitudes or lack of previous experience makes it difficult for either men or women to learn professional behaviors. The study of emotional problems which may be associated with such learning is itself an interesting research problem.

Implications for Training:

Although this research does not lead to detailed suggestions for specific modifications and developments in teacher training programs, the results are sufficiently solid to justify some general recommendations. First, it is clear that the general strategy underlying our experimentation provides a model for organizing training programs. We have been successful in teaching specific skills, although we have yet to demonstrate their general validity for producing pupil changes.

However, our practical experience is that these skills generally make sense to secondary teacher trainees, are relatively easy to teach, form the basis for extensive discussion on teaching problems, and are useful guides for continued supervision of trainees. We reiterate that the data supporting this statement is drawn from reflecting on our experience.

We experienced relatively little difficulty in engaging the Stanford Secondary Education interns in these experiments. Such difficulties as we had were the results of failures on our part to provide adequate information, or misunderstandings that were easily remedied in successive experiments. The trainees generally liked the procedures, we had almost unanimous acceptance of the videotape practices employed, and we felt that the trainees were pointed to a serious consideration of the elements of effective teaching.

No great difficulty should be experienced in combining these skills, if one wishes to train for more complex behaviors. Similarly, the kinds of decisions that the use of these skills requires is fairly obvious. Discussing these decisions would broaden any program built upon a system of skill training.

Our research does support the general rule that the use of feedback procedures, comprised of self-viewing of one's videotaped teaching performance while being assisted by a supervisor (the experimenter in these studies) is a highly effective way of modifying teaching behavior. The results of this training procedure were so systematically consistent that one could proceed with reasonable confidence that it would usually be effective and, in many cases, highly effective.

We are less clear about recommending modeling procedures, although we are confident that in the long run such training procedures will prove to be effective. As noted above, the problem is to identify those kinds of teaching skills for which modeling procedures are most effective. Until progress is made along this line of research, we cannot make specific suggestions about modeling techniques that might be used in teacher training programs.

There is one exception, however, to this disclaimer. It seems reasonably clear, as an inference from the data of these experiments and from an analysis of our experience, that modeling procedures, however they may be constructed, are likely to be highly effective only when a supervisor (again, the experimenter in these studies), cues the trainee on the desired behavior to be learned from observing the model.

We are not claiming to have proved that the strategy used in conducting these experiments is the most effective way to train teachers. Certainly we have run no comparative programs to test such a conclusion. However, a moment's consideration will impress one with the large change that occurs in a teacher training program when experiments of this kind are introduced into it; when, in fact, they become a primary vehicle for a portion of the training. Having produced this change with reasonable acceptance on the part of out trainees, we are confident that this experience points to the feasibility of making such training procedures a major part of a teacher training program.

Many problems remain to be resolved in utilizing such procedures and training programs. We have only preliminary information on the transfer effect of this training, information, however, which suggests a positive transfer effect. But our most powerful transfer studies remain to be done. We have not solved how one integrates training in decision-making with training on teaching skills simply because we have not had time to devote our attention to that complex problem. Nor have we yet demonstrated that these teaching skills effectively change pupil behavior in significant ways. We began simply: to demonstrate that teaching behavior could be learned if the conditions of learning were carefully controlled. We ourselves are engaged in continued research in the problems cited here. We trust that our colleagues will find that this initial foray into the analysis of such complex behaviors as teaching will stimulate further research in this domain.

Appendix A

DIRECTIONS TO EXPERIMENTAL GROUPS

EXPERIMENT I

Directions for Control Group (1)

As you view your own performance on this and the next three videotapes, try to determine your effectiveness in relation to:

(1) Indications of the aims of the lesson.

(2) Use of examples.

(3) Effectiveness of teacher questions.

(4) Amount and spread of participation.

(5) Pacing the lesson.

(6) Teacher-student rapport.

Directions given to Groups 2, 3, & 4.

PLEASE DO NOT DISCUSS THIS MATERIAL WITH OTHER INTERN TEACHERS UNTIL THE TAPING SESSIONS ARE COMPLETED IN DECEMBER.

(Early in the Winter Quarter, the experiment, its purposes and results, will be fully discussed with you).

One important aspect of increasing your effectiveness in the classroom is concerned with achieving optimal student participation in the lesson. In this and the next three videotapes of your classroom performance, we would like you to try to increase student participation in your lessons.

One way of achieving this goal is to reward student responses when they occur. In general, a student will begin to respond more frequently if he is rewarded for doing so. If his responses are ignored or punished, participation will tend to drop off.

Viewing Session One: In viewing your first tape, you are asked to:

(1) Select five (5) students who in this lesson participated the least. To do this, use the attached analysis chart and put a check mark beside the student's name every time he participates. After the tape is finished, select the five low participators. Since there may be more than 5 students who did not participate at all, select the five on the basis of alphabetical order.

Your objective over the next three taping sessions will be to try to get these five students to participate more frequently.

(2) Secondly, to help you analyze your own teacher behavior, you are also asked to:

(a) count the number of times you rewarded students when they participated.

(b) count the number of times you punished such responses.

(c) count the number of times you ignored such responses.

NOTE: DO THIS ANALYSIS OF YOUR RESPONSES TO THE STUDENTS FOR EACH OF THE FOUR VIDEOTAPES (analysis charts are attached).

A Student Participatory Response: (for your purposes of analysis) is any verbal response. Do not count gestures as this would overly complicate your analysis. Once a student has responded, determine your reaction to his behavior, and classify it as rewarding, ignoring, or punitive.

Teacher Responses: In classifying your responses to student participation you should look for clear-cut responses. What your intentions are, and how you feel about a student are not relevant when you chart your reactions to his participatory behavior. Look for clear-cut teacher behavior, and classify your behavior as rewarding, ignoring or punitive. Do not classify your response on the basis of how you feel about a particular student.

Scoring charts are attached to help you analyze your reactions to student participation.

Appendix B

RATER CODE SHEET
EXPERIMENT I

FORM II

Tape Code No. ____________________ Name of Intern ____________________

Print

Tape Segment ____________________ Intern Code No. ____________________

Number of Rater ____________________ Sex of Intern ____________________

Date of Rating ____________________ Tape Session ____________________

Grade Level ____________________ Date of Taping ____________________

Rater Evaluation of Tape Quality: Sound __________ Video __________

	Type of Lesson Time of Lesson	Sex of Student	Student Number	Direct Solicit or Volunteer	Timing of Reinforcement T Affect Loading of Reinforcement	T Non-Verbal Reinforce. Positive and Negative	T Verbal Reinforcement Positive and Negative	Teacher Response Style	Student Response
	Type and Time	M or F	No.	DSI DSI-I (v) V, v? (/)	B.D.A. -Ot	N.S. F Off	OK,VG,Gp,F Uh. D,Q CA (+or-) N.R., N.C., S,V. (1) X	Rd,Rf,Rp Rs,Ci,Su TF,+Q -Q PR,Rp TC, BB	OG,OI, Ch IO I- Lt,LP
	1	2	3	4	5	6	7	8	9
L									
G.W.									
I.S.									
Total									

Column	Date Information: Sheet One (Pencil)	Expected Symbols
1	Tape code number (1 - 210)	
2	Tape segment letter (A - D)	
3	Treatment code number (1 to 4)	
** 4	Experimental session number (1 to 4)	
5	Sex of Intern	
** 6	Age of Intern	
** 7	Grade level taught	
8	Date of taping	
** 9	Date of playback	
**10	Difference score (3 scores here)	
**11	School lesson taped at	
**12	Home state of Intern	
13	Tape sound quality RATER (-3 to +3)	
14	Tape picture quality RATER (-3 to +3)	
*15	Tape sound quality INTERN (-3 to +3)	
*16	Tape picture quality INTERN (-3 to +3)	
17	Raters'names (coded by last names)	
18	Date of rating	
*19	Questionnaire #1 - increased teaching effectiveness (-3 to +3)	
*20	Questionnaire #2 - tapes helpful (-3 to +3)	
*21	Questionnaire #3 - supervision helpful (-3 to +3)	
*22	Questionnaire #4 - viewing becomes boring (-3 to +3)	
*23	Questionnaire #12 - 20 minute sessions too long (yes or no)	
*24	Questionnaire #12 - 20 minute sessions just right (yes or no)	
*25	Questionnaire #12 - 20 minute sessions too short (yes or no)	
*26	Questionnaire #13 - Useful to see others' tapes (yes or no)	
27	Tape time lecture-discussion (LD) + minutes & tenths of minutes	
28	Tape time Group Work (GW)+ minutes & tenths of minutes	
29	Tape time Individual Study (IS) + minutes and tenths of minutes: T reads to class and CO	
	COLUMN 2 & 3	
30	Total number of male responders	
31	Total number female responders	
32	Total number male responses	
33	Total number female responses	
34	Total number of slashes	
35	Total number of responses (inculdes slashes): Add Col. 32, 33, 34	
	COLUMN 4	
36	Number of D S I male	
37	Number of D S I female	
38	Number of D S I total	
39	Number of D S G total	
40	Number of D S G-I male	
41	Number of D S G-I female	

Column	Date Information: Sheet One (Pencil)	Expected Symbols
42	Number of D S G-I total	
43	Number of D S G-V male, also D S G-V-I	
44	Number of D S G-V female, also D S G-V-I	
45	Number of D S G-V total, also D S G-V-I, also D S G-V where no M or F indicated	
46	Number of D S S total	
47	Number of V male	
48	Number of V female	
49	Number of V total, also V's where no M or F indicated	
50	Number of V? male	
51	Number of V? female	
52	Number of V? total, also V?'s where no M or F indicated	
	COLUMN 5	
53	Time of reinforcement (B) male	
54	Time of reinforcement (B) female	
55	Time of reinforcement (B) total	
56	Time of reinforcement (D) male	
57	Time of reinforcement (D) female	
58	Time of reinforcement (D) total	
59	Time of reinforcement (A) male	
60	Time of reinforcement (A) female	
61	Time of reinforcement (A) total	
62	Number of (+) affect loading male	
63	Number of (+) affect loading female	
64	Number of (+) affect loading total	
65	Number of 0 affect loading male; Add col. 53, 56, 59, subtract col. 62	
66	Number of 0 affect loading female: Add col. 54, 57, 60, subtract col. 63	
67	Number of 0 affect loading total: Add col 65 & 66	

Column	Data Information: Sheet Two (Red)	Expected Symbols
	COLUMN 5	
1	Number of Nods and Smiles (N S) total, or N or S	
	COLUMN 6	
2	Number of Moves toward (←) total	
3	Number of Moves away (→) total	
4	Number of Eyes off (OFF) total	
5	Number of Frowns (F) total	
	COLUMN 7	
6	Number of statements including word Good total	
7	Number of right (R) total, also T.R., Y.R., (Not A.R.)	
8	Number of fine (F) total, also T.F.	
9	Number of excellent (E) total, also Ex.	
10	Number of Okay (OK) total, also all right A.R.	
11	Number of Um-Hmm (UH) total	
12	Number of Yah or Yes (Y) total	
13	Number of +Names or +Call's attention total, (+N, +Ca, +Na)	
14	Other positive verbal reinforcement total, i.e., T.Y., I.P., W., etc.	
15	Verbal overflow (+) statements total	
16	No. wrong (X) total	
17	Negative naming (-Ca, -N) total	
18	All negative statements total, i.e., (T-), (B-), (PP-), (NR-), (S-), etc.	
19	Verbal negative overflow (-) total	
20	Grand total positive non-verbal reinforcement, add col. 1,2, (red)	
21	Grand total negative non-verbal reinforcement, add col. 3,4,5.	
22	Grand total positive verbal reinforcement, add col. 6 to 15	
23	Grand total negative verbal reinforcement, add col. 16 to 19 (or 16, 17, 18 and 19)	
24	Teacher No response (NR or NC) males	
	COLUMN 8	
25	Teacher No response (NR or NC) females	
26	Teacher No response (NR or NC) total	
27	Teacher Response Style Redirect (Rd) total	
28	Teacher Response Style Refocus (Rf) total	
29	Teacher Response Style Rephrase (Rp) total	
30	Teacher Response Style Repeat (R2) total	
31	Teacher Response Style Clarification (Cl) total	
32	Teacher Response Style Summarize (Su) total	
33	Teacher Response Style T feedback (TF) total	

Column	Data Information: Sheet Two (Red)	Expected Symbols
	COLUMN 8	
34	Teacher Response Style Qualified Reinforcement (QR, or +Q) total	
35	Teacher Response Style Qualified Reinforcement (-QR, or -Q) total	
36	Teacher Response Style Post-Hoc Reinforcement (PR) total	
37	Teacher Response Style Role Play (RP) total	
38	Teacher Response Style T clarifies (TC) total	
39	Teacher Response Style Blackboard (+BB) total	
40	Teacher Response Style Blackboard (-BB) total	
	COLUMN 9	
41	Student Responses No Group Response (OG) total	
42	Student Responses No Individual Response (OI) total	
43	Student Responses Chorus answer (Ch) total	
44	Student Responses Irrelevant Response (IO) total	
45	Student Responses Interrupt Negative (I-) total	
46	Student Responses laugh at T (LT) total	
47	Student Responses laugh at S (LP) total	
**48	Subject matter major of intern	

Appendix D

RATING DIRECTIONS

EXPERIMENT I

I. General Information:

Tape Session ____________
Intern Code No. __________
(Rater leave this blank)

____

1. Name of Intern ______________________(PRINT last name)
2. Name of Rater ______________________Date of Rating ____________
3. Sex of Intern (circle one) M F
4. Tape Code: (AA to ZZ) ____________(on tape box)
5. Tape Segment (circle one): 1st 2nd 3rd 4th
6. Rater evaluation of tape quality:

+3:	I agree very much	-3:	I disagree very much
+2:	I agree on the whole	-2:	I disagree on the whole
+1:	I agree a little	-1:	I disagree a little

a. The sound quality of the tapes was good enough to permit me to adequately analyze the performance of the intern: ___________

b. The video (picture) quality of the tape was good enough to permit me to adequately analyze the performance of the intern. _____

II. Tape Analysis

Column I: Type of lesson and Time: (1) In most cases, rate this at the end of the lesson. Most lessons will be predominately lecture and discussion (L.D.), and will run for 20 minutes. Thus, code them as L.D.-20. (2) In some cases teachers will break the class up into small groups for project and small group discussion for part or all of the tape. When this happens, record as G.W. (for group work), and indicate time spent, e.g., G.W.-10. When you have both L.D. and G.W., code them both and indicate exact time. e.g., G.W.-11, L.D.-9. (3) In a few cases there will be desk work such as tests, reading, homework, where no T-P interaction occurs. Code this as I.S., and indicate time, e.g., Teacher spends first 5 minutes of tape-time giving P's a reading assignment and then discussing the assignment with the class. Code as I.S.-5, L.D.-15.

Summary: L.D.: Lecture/discussion
G.W.: Group Work
I.S.: Individual study

NOTE: Run each tape through to the finish and record exact time -- not all tapes are 20 minutes.

Column 2: Student Number: The purpose of this column is to find out how many student responses (total number) occur during the tape. A response is any uninterrupted vocal response by a student. If a student and teacher have a verbal exchange, his responses will be recorded as follows:

S:	"What time is it?"	1
T:	"5:00 P.M.."	
S:	"Are you sure?"	1
T:	"Yes."	
S:	"That is late."	1

On the example, 3 student responses have occurred, and all of them were made by student number one. Scored this way we can get total number of S responses, and how many times each S responds by adding up the column of figures.

Column 3: Pupil's Sex: In all but a few cases you can identify this easily. Score as M or F.

Column 4: Direct Solicit or Volunteer (D.S. or V.): In direct solicit, T asks P a question. In volunteer, P volunteers a comment without being asked by T. On direct solicit by T, code this as:

D.S. I: D.S. of individual pupil
D.S. G: D.S. of group response

Column 5: Pupil No Response: T asks P a question and he does not answer. Code this simply by a stroke (/). If T asks a general question of the class, and there is no response, code it as (C).

Column 6: Timing of T Reinforcement:

1. T reinforces S before he begins to speak (B)
2. T reinforces S during his speech (D)
3. T reinforces S after his speech (A)
4. Uncertain or unknown (U)

Column 7: Affect Loading of T Reinforcement: The word "good" (or any other reinforcement) can be said by T in a negative, neutral or positive manner. If T reinforces with enthusiasm, the affect loading (emotional overtone) is positive, so code it (+). If neutral code it (0). If it is obviously negative, punitive or sarcastic, score it (-).

Column 8: T Verbal POSITIVE Reinforcement: T uses words such as:

1. Good, or very good: (V.G.)
2. Good point: (G.P.)
3. Interesting point (I.P.)
4. Good question: (G.Q.)
5. Right: (R)
6. Okay: (O.K.)
7. Fine: (F)
8. Umm-hmm: (U.H.)
9. Good beginning: (G.B.)
10. Calls class attention to S positively: (C.A.); e.g., "Listen! John has a good point."

In general, use the first letter of the word in coding. You will tend to be getting combinations such as T says "good", then nods and smiles; code this as G., N.S.. T may say, "Good, fine. That's fine! It was a good question." Code this as G., F., G.Q..

NOTE: If this proves to be impossible in actual rating, we will simply use a stroke (/) to set the number of such terms.

Column 9: Urges and Encourages: T says to S during S's answer such things as: "Yes", "Go on", "Um-hm", etc.. Use a stroke (/) for each T response unit.

Column 10: T Negative Verbal Reinforcement: Here T will:

1. Threaten Student (T) "If you do that again....."
2. Belittle Student (B)
3. Recall past punishments (RP): "Remember that once before....."

4. Require task to be done over (DO)
5. Impose physical restriction (PR): "Sit down and shut up!"
6. Send student out of room (O): "Go to the office!"
7. Withdraw privileges (WP): "No free periods because you were bad."
8. Recall past instructions negatively (NR): "Had you paid attention you would have known that.:
9. Call attention to inappropriate behavior (CA): Here T focuses class on, for example, John's "Bad" behavior.
10. Give negative command (NC): "Get to work and be quiet."
11. "Name" the student in a warning and negative manner (Ns): "John! .."

Column 11: T NON-VERBAL Positive Reinforcement: Instead of (and in combination with) + words, T may:

1. Nod and smile (N & S)

2. T moves towards S (←)

3. T keeps eyes on S (on) as S is answering or going to answer a question.

Column 12: T NON-VERBAL Negative Reinforcement:

1. T scowls, frowns (F)

2. T moves away from S (→)

3. T takes eyes off S while S is speaking. (off)

Column 13: T Informative Feedback: In response to a S's question or comment, T gives the answer without commenting on S's comment.

e.g., S: "What is the time?"
T: "3 P.M."

NOTE: Code T response of "No," "Wrong," "Incorrect" here as: X

Column 14: T No Response: T does not respond to S but makes no comment or changes the subject, goes back to the board, etc. Code by stroke (/).

Column 15: T Repeat: T repeats S's answer:

1. R O - T repeats S's answer only -- does not comment on it.

2. R + - T repeats S's answer and says "good", for example. This would be coded R-G

Column 16: T Techniques: Redirects (Re): Here T changes direction of participation by turning to another P and saying:

1. Redirects
    1) "Do you agree with that?"
    2) "Can you add to that?"
    3) "Can you help him (us) out here?"
    4) "How do you feel about that?"
2. Teacher Qualified Reinforcement:
    1) Positive (+0): Here T tells P he is partly correct, then qualifies: e.g ,
        a. "Good as far as you have gone, but . . . "
        b. "You are on the right track, but . . . "
        c. "Fine, however . . . "
        d. "Right, but don't forget . . ."
    2) Negative (-0): as above, except negative.
        a. "That's foolish, you forget . . ."
        b. "Of course not. Anyone knows that . . . "
        c. "No! Don't you know that . . . "
3. Post hoc Pupil Reinforcement (PR): T, while discussing a point says such things as: "John made a good point on this earlier."
4. Role Play (RP): T asks class to "put yourself" in the situation; e.g., "Suppose you were a fireman ... on a desert island . . . in charge of the country . . . etc."
5. Chorus Answers: the Class answers in chorus (Ch)

Appendix E

INTERN QUESTIONNAIRE RESULTS

EXPERIMENT I

Question One: "The opportunity to view the four tapes of my own teaching increased my effectiveness in the classroom."

Question Two: "The supervision I received during the tapes was helpful in increasing my classroom effectiveness."

Question Three: "In general, viewing the four tapes was an enjoyable experience for me."

Question Four: "After some time viewing one's own performance by videotape becomes boring."

	N*	Question 1	Question 2	Question 3	Question 4
I. Controls	15	1.93	-1.92	2.00	0
II. Self-Reinforcement	13	1.15	-1.46	1.92	+0.31
III. Minimal Supervision	14	1.78	+1.35	1.57	-0.14
IV. Maximal Supervision	12	1.91	+2.66	1.91	-2.00

*note: Did not get questionnaire results from all S's in the study. Complete sampling would have yielded 18 from each experimental group.

Question Seven: "We would appreciate your criticisms and comments on the taping sessions."

Question Eight: "We would appreciate your suggestions for improving future taping sessions similar to those just carried out."

Question Nine: "We would appreciate your suggestions for making supervision more effective."

The following thematic breakdown is for group 4 (maximum supervision). Questions seven and nine are combined since there was extensive overlap in the content of intern responses.

Summary of questions 7 and 9 for group 4:	Frequency of Mention
1. More time should be allowed for each playback to permit more discussion between intern and supervisor.	1

Summary of questions 7 and 9 for Group 4 (Cont'd)	Frequency of Mention
2. More sessions like those in the experiment are needed.	3
3. The supervision I received during the session has been the most valuable feature of the intern program to date	4
4. The supervision during the sessions was very helpful (and/or extremely valuable).	6
5. The supervision was generally helpful.	1
6. Supervisor approach was "a bit pressured and bull-dozing."	1
7. Supervisor's remarks were "too planned" so need more time in playbacks.	1
8. Supervisor should not censure the intern.	1
Summary of Question 8 (Group 4)	
1. More advanced warning as to when we will be taped.	5
2. Sessions should be more spread out over time.	2
3. All interns should receive this supervision.	2
4. Supervision should be integrated into the intern program more fully, and Stanford supervisor and E work together.	1
Summary of Questions 7 and 9 (Group 3)	
1. Supervision was most (very) helpful.	5
2. Supervision was helpful (worthwhile).	3
3. More supervision and guidance needed of broader scope and more extensive.	6
4. Would prefer experienced teacher rather than graduate student in psychology for supervision	1
Summary of Question 8 (Group 3)	
1. Pupils were distracted by taping sessions.	3
2. There were too many tapes.	2
3. More tapes are needed.	3
4. More time in playbacks needed.	2
5. Integrate the sessions into the intern program.	2
6. More advanced warning as to when we will be taped.	3

Summary of Question 8 (Group 3) Cont'd	Frequency of Mention
7. Sessions should be more spread out over time	4
8. When found out it was an experiment was very displeased.	1

Appendix F

INTERN INFORMATION

EXPERIMENT II

During the fall we were trying out various patterns of supervision to determine how best to use supervision with TV recording. This series we called Experiment #1. Some of you received additional supervision and some did not. Now we are changing the variable we are working with, and everyone will receive the same amount and type of supervision. We will be varying the time of taping and playbacks instead of style of supervision.

In the fall the goal was to increase the amount of pupil participation. A full report on results is forthcoming. The results are now being analyzed. This spring, in the recording series which we are calling Experiment #2, you will have an opportunity to develop specific skills that will help you upgrade the quality of student responses. The intern or teacher behavior of interest in this experiment is redirecting and refocussing pupils' questions and statements to prompt their self-correction and renewed attempts at answers. This is a useful skill for the intern to acquire, for it allows him to control classroom interaction in such a way that pupils may develop increased critical self-awareness, more clearly analyze the implications of their own and others' questions and conclusions, facilitate the transition from relatively superficial to more complex levels of analysis of a problem, and to facilitate transfer.

These in general are to be your goals during the coming weeks as you are videotaped. The videotape provides all of us with an excellent means for practicing and mastering these skills. Some minor restrictions on your lessons and time for playbacks are more than offset by the potential gains in professional skill.

The objective of the current study is to help interns develop more alternatives in responding to student questions and statements.

BASIC RULE I: The Teacher (T) does not give immediate answers to pupils' questions.

Specific Teacher Techniques (where appropriate)

1. Ask the pupil (P) to break his question down into more easily answered parts.
2. T rephrases P question without giving the answer (this is a technique developed by Carl Rogers-- the idea here is to rephrase the question in such a way as to elicit further clarification by P).
3. T may say, "What things (or aspects of the problem) must we consider in order to get the solution; i.e., where do we look for the answer?"
4. T redirects the P's question or statement to another P or P's by saying such things as:
   (a) "Can you help him out?"
   (b) "Do you agree with that?"
   (c) "Can you add to that?"
   (d) "What are the implications of this for . . .?"
   (e) "Can you summarize this for us?"
5. Role Play: P may ask what the reasons are for doing something in a particular way (e.g., P asks T, "Why did Germany attack Poland before France?"). T's response here is to say, "All right, I want you to be that person (or in the situation for a moment. How do you justify the action (method) you have taken?"

6. T Qualified Response: Assuming P has no idea of the answer (after probing by T), T may give part of the answer to P and require him to find a solution and report back to class (or a legitimate question for group work later on).

7. Encouragement of Alternatives: In response to a P question or statement T may say, "Before we try to reach a conclusion (or answer) here, let us explore as many avenues as possible. Think of as many (answers, reasons, solutions) as you can." Encourage your students to divergent thinking.

BASIC RULE II: T does not immediately accept P's answer or statement.

Possible T techniques:

1. Clarification: T asks P to rephrase or clarify his answer or statement.

2. Summarize: T asks P to summarize his answer, statement, or class discussion.

3. Critical Awareness: T asks P (or class in general) such things as:

   a) "What are we assuming here?"
   b) "Is this one question or several questions?"
   c) "Can we break the problem down into more manageable parts for analysis?"
   d) "Have we (you) oversimplified this -- is there more to it?"
   e) "How would someone who took the opposite view respond to this?"

4. Redirect: Here, as outlined earlier, T changes the direction of participation to another pupil and says:

   a) "Do you agree?"
   b) " . . . add to that?"
   c) "What else might be said?"

5. Refocus: T refocusses discussion or class attention on another question or problem (usually a related one, but not necessarily so) -- T may say:

a) "Good! What are the implications of this for . . .?"
b) "Let's look at it this way . . .; what is your answer now?"
c) "How does this relate to . . . ?"

As you can see, these specific techniques which are appropriate for various situations are inappropriate in others. In preparing your lessons for TV, do all you can to allow yourself the opportunity to practice these techniques with the class. In the playback sessions, we will provide additional instruction and specifically relate the above-named and other techniques to your particular situation. In general, discussion-type lessons, with material that allows you and the class to interact and analyze problem areas will be best. Please avoid individual seat work, tests, quizzes, and group work on taping days.

BASIC RULE III: As the students start moving toward the goals you have set, reward these desirable responses. At first their attempts may be far short of the ideal; to get them to approach that ideal, reinforce or reward them at each step along the way. Don't wait for the perfect answer -- reward each effort so that they move forward bit by bit.

GENERAL INFORMATION:

1. Interns will generally be given at least one week's notice prior to taping dates. We are anxious to avoid the mistakes we made in this regard in the fall study. There are some situations which are beyond our control - we ask that you recognize the possibility of (a) machine breakdowns (b) faulty videotapes (c) interns missing playbacks. If this happens, you will be notified promptly and the taping will be rescheduled. Most of these bugs have been ironed out so that we really don't expect holdups of this kind (we have more machines now so scheduling is facilitated).

2. As usual, you will be taped for 20 minutes on each of four lessons. You will be asked to see these tapes in playback sessions.

Appendix G

DIRECTIONS FOR INTERN TEACHERS

EXPERIMENT III

Following the pretest, each subject received these directions in Step Two of the general procedure. The two Symbolic Modeling groups reread the instructions in Steps Four and Eight.

DIRECTIONS FOR INTERNS: BASIC QUESTIONING TECHNIQUES

READ THE FOLLOWING INSTRUCTIONS VERY CAREFULLY BEFORE VIEWING THE VIDEOTAPE

Today you will have an opportunity to develop skills in basic classroom questioning techniques. The session is designed to help you extend the range and quality of your questioning techniques in such a way that the pupils you teach are led to think more deeply about problems raised in class.

The techniques outlined below are designed to be used in discussion, review, and inductively organized lessons where active pupil participation is prerequisite to the realization of the goals of instruction. Any given technique may be appropriate in one situation but not in another. The selection of a particular technique depends upon the extent to which, in your judgment, it requires the pupil to analyze critically a problem or justify rationally his answer. Do not use a given technique unless you feel it contributes to the educational relevance of the lesson.

YOUR GOAL IS TO ASK PENETRATING AND PROBING QUESTIONS THAT REQUIRE PUPILS TO GO BEYOND SUPERFICIAL, "FIRST-ANSWER" RESPONSES.

BASIC QUESTIONING TECHNIQUES: There are two ways of achieving the above goal: 1) The teacher asks penetrating questions that require pupils to get the heart of the problem. This forestalls superficial answers. Whether or not you are able to do this depends largely upon your knowledge of relevant content. 2) The second approach is based on specific techniques that may be used after the pupil has responded in some way (i.e., a question, a comment, an answer to a teacher question). The goal here is to get the pupil to go beyond his first response. You attempting to produce greater critical awareness and depth by Probing. Your cue is the pupil's response--once it has occurred, don't immediately go on with the discussion yourself. Probe his answer by means of one of the techniques outlined below.

I. Teacher Seeks Further Clarification by the Pupil: You may ask the pupil for more information and/or more meaning. You may respond to the pupil's responses by saying such things as:

a) "What do you mean? "

b) "Please rephrase/clarify what you mean."

c) "Can you explain that further?"

d) "What do you mean by the term . . .?"

II. Teacher Seeks Increased Pupil Critical Awareness: Here you are requiring the pupil to justify rationally his response. You may say:

a) "What are you/we assuming here?"
b) "Why do you think that is so?"
c) "Have we/you oversimplified the issue--is there more to it?"
d) "Is this one or several questions?"
e) "How would someone who took the opposite point of view respond to this?"

III. Teacher Seeks to Refocus the Pupil Response:If a pupil has given a high quality answer, it might seem unnecessary to Probe it. However, you could Refocus his or the class's attention on a related issue.

a) "Good! What are the implications of this for . . .?"
b) "How does this relate to . . .?"

c) "Can you take it from there and tie it in to . . .?"

IV. *Teacher Prompts Pupil*: In *Prompting*, you are giving the pupil a *hint* to help him go on and answer a question. Supposing a pupil has given an, "I don't know", or, "I'm not sure" type of response. Rather than giving him the answer or redirecting the question to another pupil, you may give the puzzled student a hint.

e.g., Teacher: "John, define the term polygenesis."

John: "I can't do it."

Teacher: (prompt) "What does *poly* mean?" or, "Well,
means origin or birth, and *poly* means . . ?"

This technique allows you to Probe even though at first, it appears that the pupil can't answer the question.

V. *Redirect*: This is not *Probing* technique *per se*. It helps you bring other students into the discussion quickly while still using *Probing* techniques. In *Redirect*, you merely change the direction of interaction from yourself and pupil (one) to yourself and another pupil (two).

e.g., Teacher: "What is the relationship between pressure and volume?"

p^1: "As pressure goes up, the gas is condensed."

Teacher: (to p^2) "Can you tell us what is meant by condensed?" Or, "Can you restate that in terms of volume, p^2?"

To sum up, the techniques outlined above have two things in common:

1) They are initiated by the teacher immediately after the

pupil has responded.

2) They require the pupil to go beyond the information he has already given.

CONCLUDING REMARKS: Try to use the techniques as frequently as you can. Do not stay with one given technique for too long at one time. In addition, don't forget to reinforce when you Probe--there might be a tendency at first to behave like a "Philadelphia lawyer."

If you prefer to run through the first five-minute lesson as a warmup, this would be fine. You may teach the same lesson over two or three times. We will focus more on Probing than on transmitting new or complex material. The maximum amount of time for the session will be two hours.

You will be working with ninth and tenth grade pupils who have had some experience with modeling sessions and who will be very cooperative and responsive.

Thank you for your help.

FOR TRAINEES IN EXPERIMENTAL CONDITIONS USING VIDEO-TAPED TEACHING PERFORMANCES OF THEMSELVES OR OF MODELS

The following instructions were added for these trainees:

With these ideas in mind, you may now critically view your own performance. Other parts or segments of the session will give you an opportunity to practice and reflect upon these techniques.

One last note: In addition to giving you training in basic questioning techniques, we are trying to determine the best way to teach them. We are using several training methods. All of them have been shown to be effective, but we want to find out which particular method is most effective. Thus, different groups of teachers will be taught the same thing in different ways.

Please do not discuss any part of your training with other teachers until the quarter is over.

We will review the techniques and give you the results once the data are analyzed.

Appendix H

THE RATER'S MANUAL

EXPERIMENT III

THE RATER'S MANUAL

The Rater's Manual served as the basic training device in developing rating skills. It will be noted that the manual is keyed to The Rater's Code Sheet. The latter was used to record the relevant behaviors as the tape was viewed by the rater. The reader may find it helpful to refer to the Code Sheet before reading the manual.

Note that Columns (1), (3), (13), (14) and (15) are not discussed in the Directions for Interns. Column (1) served a dual function. First, it was found that by requiring the raters to record these behaviors, reliability was increased. Raters were forced to make finer distinctions between Prompting and non-pupil-response contingent teacher verbal behavior. In addition, Columns (1) and (14) served to provide an index of the generalization of Probing techniques to other areas of teacher verbal behavior.

Column (3) was included to see whether or not the intern's rate of reinforcement dropped off with an increase in Probing. The two techniques are deemed by the investigator to be complimentary rather than discrete strategies. In fact, Probing grew out of an earlier investigation where reinforcement was used to increase pupil participation. Probing was developed to increase the quality of such participation.

Columns (13) and (15) constitute relatively low-frequency Probing techniques. They were not included in the written instructions to the interns as it was felt that a plethora of specifics would tend to increase interference. The models were, however, instructed to demonstrate both of these techniques at least once in their tapes.

The use of each of the latter techniques was pointed out to the intern in discrimination training phases of the experiment. However, the experimenters did not single them out for special attention.

THE RATER'S MANUAL

GENERAL INFORMATION:

a) Tape Box Number __________

b) Segment rated from __ to __

c) Rater Evaluation of Tape Quality:

   a. sound ________

   b. video ________

e) Name of Intern ______________

f) Sex of Intern ______________

g) Tape Session No. ____________

h) Date of Taping ______________

i) Subject Matter of Lesson ______________

d) Name of Rater and date ____________________

   Type of Lesson and Time:

   L.D. ___________________

   I.S. ___________________

   G.W. ___________________

Raters are responsible for filling in all of the above data. The bulk of the information is to be gathered after the tape has been rated. Print each bit of information as clarity is essential.

a) Tape Box Number: Ask operator for this number before you begin to rate, and before it is put on the machine, if possible.

b) Segment Rated: Get precise beginning and ending times (accurate to 1/10 of a minute) of the segment of tape rated. This should be read off the machine by the operator and checked also against the information on the box.

c) Rater Evaluation of Tape Quality:

+3 - I agree very much	-3 - I disagree very much
+2 - I agree on the whole	-2 - I disagree on the whole
+1 - I agree a little	-1 - I disagree a little

1) The sound quality of the tapes was good enough to permit me to adequately analyze the performance of the intern: ____________

2) The video (picture) quality of the tape was good enough to permit me to adequately analyze the performance of the intern: ______________

3) Name of Intern: write in after tape is rated.
4) Sex of Intern: " " " " " "
5) Tape Session No: " " " " " "
6) Date of Taping: " " " " " "
7) Subject Matter of Lesson: " " " " " "

TYPE OF LESSON AND TIME: Rate this (as a rule) at the end of the lesson. Most lessons will be five-minute Lecture-Discussion (L.D.) lessons. Code them on the L.D. line as 5.0. In some cases, I may have group work (G.W.) or individual study (I.S.) for part or all of the lesson. If this happens, code appropriately.

Lecture-Discussion: (L.D.).____________________
Group Work: (G.S.).____________________
Individual Study (a test, silent reading, etc.): (I.S.).____________________

Be sure to run each tape through to finish and record exact time.

INTRODUCTION TO RATING: The code sheet is based on a "typical" unit of classroom interaction. Usually: (1) T asks a question.
then: (2) P responds.
then: (3) T Probes or does not Probe this pupil response. Our primary interest is in T's response. Does T Probe or Not Probe? Every unit of T-P interaction is to be coded this way.

COLUMN ONE: T REPHRASES OR CLARIFIES HIS OWN RESPONSE: While T's goal is to get pupils to Clarify and Rephrase their responses, this may spread to the teacher's responses as well. In asking questions, T may Clarify or Rephrase it, e.g.,

T: "What is ironical about this situation?"
T: "That is, are our expectations about what should happen reversed here?"

In the above example T asks a question, then clarifies it by becoming more precise in his definition of "ironical." Frequently, T rephrases on a more obvious level, e.g.,

T: "Define 'economic'."

T: "What do we mean by economics?"

It will be noted that column (1) (T clarifies or rephrases own response) is similar in some ways to column (11), Prompting. The key difference is that in Prompting, T is responding to a pupil response, whereas in column (1), he is responding to his own prior response with no intervening pupil response having occurred. Unless the pupil actually responds, T is classed as in column one.

Example of T rephrasing/clarifying his own response COLUMN (1)	Example of T prompting COLUMN (11)
T: "Define the word ironic.".... What do we mean when we say: 'that was ironic'?"	T: "Define the word ironic." P: "I'm not sure." T: (to P) "It has something to do with a reversal of our expectations---does that help? Can you give me an example of an ironical situation?"

COLUMN TWO: PUPIL RESPONSE: P may respond in one of five ways:

1) Comment (C)
2) Question (Q)
3) Answer (A)
4) No Response: T asks a question, P does not respond.
5) Rhetorical Comment/Question (C?) e.g., T has asked for possible uses of petroleum:

(both may	P: "Gasoline would be one use -- wouldn't it?"
be defined	P: "Wouldn't paving be another use?"
as C?)	

COLUMN THREE: T REINFORCES PUPIL: T may reward a P response by saying: "Good," "Fine," "Excellent," "Right," "Yes," "Correct," etc. If T uses verbal reinforcement as outlined above, then record this in column 3. Count each bit of reinforcement as a (1)

e.g.,	T Response	No. of Reinforcements
T:	"Good!"	1
T:	"Good! That's a good answer, John."	2
T:	"Correct. In fact, quite good. You know your stuff."	3

COLUMN FOUR: T REPEATS PUPIL RESPONSE: This column covers repetition and simple rephrase as when T does not add anything to the gist of a P response in repeating it, but does reword it slightly. Note that T may or may not Probe once he has reinforced and/or repeated a P response.

NO PROBE AND PROBE COLUMNS: These two half columns permit you to indicate whether or not T has Probed a P response. If T has accepted (or No-Probed) a P response then complete the (X) in the NO PROBE half-column. If T has Probed the P response, then leave the NO PROBE column blank and complete the (X) in the PROBE half-column. Every T response following a P response must be coded as a Probe or No-Probe response. Once coded, each type of response must be further classified as some particular type of No-Probe (see columns 5 and 6) or Probe (see columns 7 - 12).

A T response is classified as a Probe when, once the pupil has responded, T asks that pupil for further comment on his response. There are five types of T Probing behavior (see columns 7 - 11). A No-Probe T response occurs when T does not ask for more information, meaning or cognitive activity from P that is based on P's response. Instead T may:

1) ignore the response (col. 6)
2) reinforce only (col. 3)
3) repeat only (col. 5)
4) add more information himself (col. 5)
5) or simply tell P his answer is wrong (col. 6)

COLUMN FIVE: **T ANSWERS P's QUESTION OR ADDS MORE INFORMATION**

This is one type of NO-PROBE response.

e.g., (T answers question)

P: "Who discovered America?"

T: "Good question. Columbus did."

In the above example, T Reinforced and answered the question, but did not Probe.

e.g., (T adds information)

T: "Who is the president of the U. S. A.?"

P: "Mr. Johnson."

T: "Yes, he is a Texan who was vice-president under Mr. Kennedy."

In example 2, T Reinforced and added information, but did not Probe.

COLUMN SIX: **T NO-RESPONSE OR SAYS ANSWER IS WRONG:** This is a No-Probe Column.

a) T No-Response: P makes a comment or answers a question, and T does not respond to P verbally --he ignores the P response.

b) T says P's Answer is Wrong: Once P has responded, T may say: "No, you are wrong; incorrect; of course not!" When this occurs code it as (X) in column 6. Usually T will say (X) then (A) answer the question himself (in which case put a (1) in column 5 for T adds information) or (b) he may Probe by means of one of the techniques in columns 7 - 12.

COLUMN EIGHT: **T CLARIFIES:** (P is asked to clarify his answer): In Clarification, T is asking for more information (e.g., "Can you add to that?") or for more meaning (e.g., "What do you mean by _____?") T is trying to get P to extend the meaning of his response. Clarification then includes:
(a) WHAT questions
(b) HOW questions
(c) WHEN questions

Note: WHY questions are to be classified as Critical Awareness of Questions. (col. 9). In Clarification T asks P to rephrase or clarify his comment, answer or question. T may say:

(a) "What do you mean?"
(b) "Tell us more."
(c) "Please rephrase/clarify."
(d) "Explain further."
(e) "How would it be done?"
(f) "In what way?"
(g) "Do you mean...?"
(h) "Can you give an example of that?"

Three other types of Clarification may occur:

(1) T asks P to summarize/sum up/tie together/etc., the discussion up to that point. e.g.,

(a) "Can you sum up the discussion on this issue?"
(b) "What then is the basic point that has been made about (this problem)?"

(2) When P asks a question, T may respond by asking P to break the question down into parts to be considered one at a time. e.g.,

(a) "Can you break your question down into parts?"
(b) "Are you asking one or several questions here?"

T is attempting to get P to Clarify the problem or define the issue so that it may then be analyzed or answered.

(3) When T asks P (following a P question) where one might look for the answer: e.g.,

"Where would we look for the answer?"; "What sorts of things (problems, issues, areas of subject matter) would we have to look at first?"

COLUMN NINE: T CRITICAL AWARENESS: T is trying to get P to become critically aware of his assumptions, attitudes, opinions. T is therefore asking P to justify rationally his response. T may say:

a) "Can you support that?"
b) "Why do you believe that?"
c) "What are you/we assuming here?"
d) "Have you oversimplified this - is there more to it?"
e) T 'quizzical' response: e.g., T may say: "Oh?" which implicitly

asks P to justify his response. In addition T may repeat the P response as a question: i.e.,

P: "The butler did it."

T: "The butler?"

f)T may respond to a P response by saying: "Are you sure?" In effect, T is asking P to reflect on his response to make sure that he is not overlooking or leaving anything out --T is asking P to analyze critically his response.

Note: Both (e) and (f) above may appear to some extent to be classifiable as one or another of the Probing techniques. To avoid confusion, they have arbitrarily been classified as Critical Awareness examples. If we subsequently find other specific T responses that are doubtful, write them down and submit them to the Research Assistant supervising the rating. He will classify them and incorporate them into an appendix.

Remember, the basic function of a T Critical Awareness Probe to a pupil response is to get P to justify his answer. It also includes questions that ask P to identify his assumptions. "Why" questions are included in this category.

COLUMN TEN: REDIRECT: Following a P response, T may turn to another P and say:

a) "Do you agree?"

b) "Can you help him/her out?"

c) "Can you add to that?"

d) "What has he/she forgotten?" etc.

In Redirect, T merely changes the direction of participation by turning to another P. Most frequently, T will use *Redirect* in conjunction with another *Probing* technique, e.g.,

P^1 response: "Sex is funny."

T response: "Fine, can you add to that, Bill?"

In the above example, T *reinforced* P, then used *Redirect* by *Clarify* (asking for more information and meaning). In this case, score both columns. Remember *redirect* occurs immediately after the P^1 response; if it does not, then T has asked another question (direct solicit) of P^2 and has not used *redirect*.

COLUMN ELEVEN: T PROMPT: Assume that T is trying to *Probe* P, but he does not fully respond. Rather than accepting this "I don't know" type of response, T leads him by asking another question.

Column Twelve: T REFOCUS: This includes recentering (a situation where P makes an irrelevant response, and T asks P how his response is related to the issue at hand). *Refocus* is signalled by T phrases such as:

a) "What are the implications of your answer for . ..?"
b) "How does this relate to/tie into . . . ?"
c) "If we look at it this way . . . what is your answer then?"
d) "How would someone (or a specific person or group) answer this question?"
e) "How is your answer related to and different from the position of . . . . on this issue?"

Refocus will tend to occur when T wants to emphasize a particular facet of a problem, or when P has given a high quality answer, and T, wishing to *Probe*, asks P to spell out the implications of the answer for related issues. Refocus as Probe only occurs when T gets P or the class to relate, tie in, or give implications for the previous P response.

Note: The following columns (13, 14,15) relate to T techniques as opposed to T response styles as in the above columns. Thus they are separated from the rest of the code sheet by a double line.

COLUMN THIRTEEN: T ENCOURAGES ALTERNATIVES: The opposite of Encouraging Alternatives (or divergent thinking) is convergent thinking where T leads P to converge on the single, right, correct answer. In encouraging alternatives, T gets P or the class to list as many possible or probable solutions/reasons/ answers/as possible. Here speculation on a broad front is encouraged by T. It is akin to brainstorming where you get as many responses as possible, ignoring quality of response. If T stops to discuss and analyze each response by the class, then he is NOT Encouraging Alternatives. The analysis part comes after the list is on the board. Here T goes on to get the class to evaluate each of the listed possibilities.

COLUMN FOURTEEN: T SUMMARIZES FOR CLASS: Instead of asking P to sum up, T summarizes by saying in effect: "This is what we have been doing for the past few minutes."

COLUMN FIFTEEN: T USES ROLE PLAY: T asks P to put himself-herself/ themselves in someone else's shoes. The intent here is to get P to identify with the situation so as to suppress superficial answers). Thus T may say:

"Pretend you are . . . . . .What would you do/did/say, etc.?"

RATER'S CODE SHEET

Page#

SEGMENT RATED FROM: ________ TO: _____

INTERN NAME: ____________________

TAPE BOX NUMBER ____________________

INTERN SEX: ____________________

RATER NAME AND DATE ____________________

TAPE SESSION NO: ____________________

RATER EVAL. OF TAPE QUALITY: ____________

DATE OF TAPING: ____________________

SOUND: ________ VIDEO: ____________

SUBJECT MATTER: ____________________

TYPE OF LESSON AND TIME

L.D. ____________ ____________________

I.S. ________________________________

G.W. ________________________________

TOTAL: ____________________

T Rephrases or Clarifies Own Response	Pupil Response	T Reinforces Pupil	No Probe	T answers quest. or adds information	No Response or T says answer is X	Probe	T Clarifies	T Critical Awareness	T Redirect	T Prompt	T Refocus	T encourages alternatives	T summarizes for class	T uses Role Play
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)

Appendix I

MODEL'S MANUAL

EXPERIMENT III

One week prior to training and taping, all models received the following directions. They were asked to be ready to teach a five-minute discussion lesson incorporating the Probing techniques.

MODEL'S MANUAL

BASIC QUESTIONING TECHNIQUES

Orientation: The basic objective of the upcoming research is to help the intern extend the range and quality of his questioning techniques in such a way that his pupils think more deeply and broadly about the problems raised in class.

We propose to do this by means of models. We would like you to teach two or three five-minute lessons in your subject-matter area to four pupils. In each lesson your goal will be to demonstrate Probing techniques to the best of your ability. Try to include as many instances of the desired behaviors as possible. The more frequently you are able to demonstrate the techniques, the better it will be in terms of training the interns. Naturally, there are limits to this. Too high a frequency of the desired behaviors may jeopardize the educational quality of the lesson.

The following discussion of specific objectives and the actual techniques we would like you to model, are written up as they will be presented to the interns.

At this point, the instructions given on pp.188ff., Appendix G, were attached.

Appendix J

INTERN VIDEOTAPE QUESTIONNAIRE

EXPERIMENT III

INTERN VIDEOTAPE QUESTIONNAIRE

Name of Intern (Print) ______________________________

Sex of Intern ______________________________

Subject Taught ______________________________

Date ______________________________

PLEASE BE SURE TO ANSWER EVERY QUESTION.

My Supervisor today was: Mr. Orme _______

Mr. Berliner _______

Mark each statement in the right margin according to how much you agree or disagree with it. Write +1, +2, +3, or, -1, -2, -3 depending on how you feel in each case.

+1	I agree a little	-1	I disagree a little
+2	I agree on the whole	-2	I disagree on the whole
+3	I agree very much	-3	I disagree very much.

1. In general, viewing the tapes with the supervisor was an enjoyable experience for me. __________
2. I now have a clear idea of what is meant by probing a pupil response. __________
3. I feel that using probing techniques in the classroom is educationally sound. __________
4. The most effective part of the total training experience today was: ______________________________
____
5. The least effective part of the total training experience today was: ______________________________
____

Below are a number of opposite terms. Rate the supervisor with whom you worked today on each pair of terms. Place an (X) in the bracket that corresponds to your assessment of the supervisor.

Examples: Supposing your supervisor was very tall you would score him as follows:

short ( ) - ( ) - ( ) - ( ) - ( ) - ( ) - (X) tall

Be sure to respond to every set of terms.

skillful	() - () - () - () - () - () - ()	unskillful
considerate	() - () - () - () - () - () - ()	inconsiderate
cold	() - () - () - () - () - () - ()	warm
likable	() - () - () - () - () - () - ()	unlikable
irresponsible	() - () - () - () - () - () - ()	responsible
interested	() - () - () - () - () - () - ()	disinterested
authoritarian	() - () - () - () - () - () - ()	non-authoritarian
critical	() - () - () - () - () - () - ()	constructive
rewarding	() - () - () - () - () - () - ()	punishing
accepting	() - () - () - () - () - () - ()	demanding
intimidating	() - () - () - () - () - () - ()	encouraging
realistic	() - () - () - () - () - () - ()	unrealistic
flustered	() - () - () - () - () - () - ()	calm
caring	() - () - () - () - () - () - ()	uncaring
friendly	() - () - () - () - () - () - ()	unfriendly

Appendix K

STANFORD TEACHER COMPETENCE APPRAISAL GUIDE

STANFORD TEACHER COMPETENCE APPRAISAL GUIDE

Following each modeling tape, the participating pupils were asked to rate the model's performance. The data reported below are the scores for the model tapes used in the study. A short form of the Guide appears on the next page. Items twelve and thirteen were not scored by the pupils.

STANFORD APPRAISAL GUIDE SCORES FOR THE MODEL TEACHERS (a)

Model	Subject	Sex	Tape Selected	Mean Score	Mean Pupil Responses										
					1	2	3	4	5	6	7	8	9	10	11
1	S.S	M	3	58.2	5.5	5.8	5.3	5.0	5.8	5.3	5.3	5.3	5.8	5.3	5.0
2	E.	M	2	59.2	5.8	5.5	5.3	5.3	5.8	5.0	5.5	5.5	5.5	5.0	5.3
3	M.	M	3	60.0	5.0	5.0	5.5	5.3	5.8	5.5	5.8	6.0	5.8	5.8	5.0
4	Sc.	M	2	61.2	6.0	5.2	5.5	5.5	5.3	6.0	6.0	5.8	6.3	5.8	5.0
5	S.S	F	2	63.0	5.3	5.5	5.5	5.8	5.5	5.3	6.5	6.2	6.0	5.5	6.0
6	E.	F	1	62.8	5.5	5.5	5.8	5.8	5.5	5.3	5.8	6.3	6.0	5.8	5.8
7	M.	F	2	57.0	5.0	5.0	5.0	5.0	5.3	5.3	5.3	5.5	5.3	5.8	5.3
7	Sc.	F	2	57.0	5.0	5.0	5.0	5.0	5.3	5.3	5.3	5.5	5.3	5.8	5.3

(a) Four pupils rated each tape. The scores reported for each of the eleven items are thus mean scores. The total mean scores were computed from the raw scores.

The norms published for the Guide indicate that the scores may be interpreted in the manner described in Table A-1.

INTERPRETATIVE CLASSIFICATIONS FOR SCORES DERIVED FROM THE STANFORD APPRAISAL GUIDE

Per Cent Scoring This High	Score	Interpretative Category
10%	7 6	Truly Exceptional, Outstanding
15%	5	Superior
15%	4	Strong
15%	3	Average
15%	2	Below Average
30%	1	Weak
0	0	Unable to Observe

THE STANFORD TEACHER COMPETENCE APPRAISAL GUIDE

1.	Clarity of Aims	The purposes of the lesson are clear.
2.	Appropriateness of Aims	The aims are neither too easy nor too difficult for the pupils. They are appropriate, and are accepted by the pupils.
3.	Organization of the Lesson	The individual parts of the lesson are clearly related to each other in an appropriate way. The total organization facilitates what is to be learned.
4.	Selection of Content	The content is appropriate for the aims of the lesson, the level of the class, and the teaching method.

5.	Selection of Materials	The specific instructional materials and human resources used are clearly related to the content of the lesson and complement the selected method of instruction.
6.	Beginning the Lesson	Pupils come quickly to attention. They direct themselves to the tasks to be accomplished
7.	Clarity of Presentation	The content of the lesson is presented so that it is understandable to the pupils. Different points of view and specific illustrations are used when appropriate.
8.	Pacing of the Lesson	The movement from one part of the lesson to the next is governed by the pupils achievement. The teacher "stays with the class" and adjusts the tempo accordingly.
9.	Pupil Participation And Attention	The class is attentive. When appropriate the pupils actively participate in the lesson.
10.	Ending the Lesson	The lesson is ended when the pupils have achieved the aims of instruction. There is a deliberate attempt to tie together the planned and chance events of the lesson and relate them to the immediate and long range aims of instruction.
11.	Teacher-Pupil Rapport	The personal relationships between pupils and the teacher are harmonious.
12.	Variety of Evaluative procedures	The teacher devises and uses an adequate variety of procedures, both formal and informal, to evaluate progress in all of the aims of instruction.
13.	Use of Evaluation to Improve Teaching and Learning	The results of evaluation are carefully reviewed by teacher and pupils for the purpose of improving teaching and learning.

Appendix L

THE STATISTICAL NATURE OF THE DEPENDENT VARIABLE

PROBING

TABLE L-1

PROPORTIONATE CONTRIBUTIONS OF CLARIFICATION IN COMPARISON WITH THE COMBINED PROPORTIONATE CONTRIBUTIONS OF ALL OTHER RESPONSE CATEGORIES TO TOTAL PROBING, BY EXPERIMENTAL GROUP AND TREATMENT SESSION

(N = 326)

Response Category	Session	Experimental Group					
		1	2	3	4	5	6
		Proportionate Contribution to Total Probes					
Clarification	Session 1	48.74	61.34	69.12	57.41	45.83	53.00
	Session 2	50.25	53.79	60.00	49.39	47.97	53.20
	Session 3	47.83	45.63	60.00	50.20	51.97	54.84
All Other Response Categories of Probing	Session 1	51.26	38.66	30.88	42.59	54.17	47.00
	Session 2	49.75	46.21	40.00	50.61	52.03	46.80
	Session 3	52.17	54.36	40.00	49.80	48.03	45.16

TABLE L-2

RELIABILITIES OF THE DIFFERENCE SCORES USED IN THE ANALYSES OF CONVARIANCE (a)
(N = 324)

Variable	Type of Difference Score		
	$(S_3 - S_1)$	$(S_2 - S_1)$	$(S_3 - S_2)$
Total Pupil Responses	.99	.98	.98
(1) Pupil Questions	.63	.53	.58
(2) Pupil Comments	.84	.87	.85
(3) Pupil Answers	.96	.95	.92
Intern Repeats	.98	.97	.95
Total Non-Probes	.99	.98	.98
(1) Interns Answers	.79	.78	.68
(2) Intern No Response	.88	.88	.85
Total Probes	.98	.98	.97
(1) Clarification	.96	.95	.91
(2) Critical Awareness	.92	.93	.89
(3) Redirection	.94	.95	.94
(4) Prompting	.87	.89	.88
(5) Refocus	.84	.81	.80

(a) Reliability of the Difference Score = $r_{dd} = \frac{r_{xx} - r_{xy}}{1 - r_{xy}}$

where, r_{xx} = inter-rater reliability

r_{xy} = session to session correlation for scores on a given variable (McNemar, 1962, pp. 155-161).

REFERENCES

Anderson, H. H., and Brewer, H. M. Studies of teachers' classroom personalities: I. Dominative and socially integrative behavior of kindergarten teachers. Applied Psychology Monograph. Stanford: Stanford University Press, 1945, No. 6.

Allen, D. W., McDonald, F. J., and Orme, M. E. J. Effects of feedback and practice conditions on the acquisition of a teaching strategy. Paper delivered at American Educational Research Association Convention, Chicago, February 1966.

Angell, D., and Lumsdaine, A. A. Prompted plus unprompted trials versus prompted trials alone in paired-associate learning. In Lumsdaine, A. A. (Ed.), Student response in programmed instruction. Washington: National Research Council, 1961 389-399.

Angell, D , and Lumsdaine, A. A. The effects of prompting trials and partial-correlation procedures on learning by anticipation. In Coulson, J. E. (Ed.), Programmed learning and computer-based instruction. New York: Wiley, 1961, 143-147.

Bandura, A. Influence of model's reinforcement contingencies on the acquisition of imitative responses. Journal of Personality and Social Psychology, 1965, 1, 589-595.

Bandura, A. Behavioral modification through modeling procedures. In Krasner, L., and Ullman, L. P. (Eds.), Research in behavior modification. New York: Holt, Rinehart and Winston, 1964.

Bandura, A., and McDonald, F. J. The influence of social reinforcement and the behavior of models in shaping childrens' moral judgments. Journal of Abnormal and Social Psychology, 1963, 67, 274-281.

Bandura, A., and Walters, R. H. Social learning and personality development. New York: Holt, Rinehart and Winston, 1963.

Bandura, A., Ross, Dorothea, and Ross, Sheila A. Transmission of aggression through imitation of aggressive models. Journal of Abnormal and Social Psychology, 1963, 66, 3-11.

Bandura, A. Social learning through imitation. In Jones, M. R. (Ed.) Nebraska symposium on motivation. Lincoln: University of Nebraska Press, 1962, 11-269.

Cock, J.O. Processes underlying learning in a single paired-associate item. Journal of Experimental Psychology, 1958,56,455-456.

Cook, J. O. and Kendler, T.S. A theoretical model to explain some paired-associate learning data (1956). In Lumsdaine, A.A., and Glaser, R. (Ed.), Teaching machines and programmed learning. Washington: National Education Association, 1960, 602-604.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1959.

Flanders, N. A. Interactional analysis in the classroom. Ann Arbor: University of Michigan Press, 1964, (mimeo.).

Flanders, N.A. Teacher influence, pupil attitudes and achievement. Minneapolis: University of Minnesota, 1960, (mimeo.).

Gage, N. L., (Ed.), Handbook of research on teaching. Chicago: Rand, McNally, 1963.

Hilgard, E. R. Theories of learning. (2nd ed.) New York: Appleton-Century-Crofts, 1956.

Hill, W. F. Learning theory and the acquisition of values. Psychological Review, 1960, 67, 317-331.

Lumsdaine, A. A. Some theoretical and practical problems in programmed instruction. In Coulson, J. S. (Ed.), Programmed learning and computer based instruction. New York: Wiley, 1961, 134-154.

Margolius, G. J., and Sheffield, F. D. Optimum methods of combining practice with filmed demonstration in teaching complex response sequences: Serial learning of a mechanical assembly task. In Lumsdaine, A.A. (Ed.), Student response in programmed instruction. Washington: National Research Council, 1961.

McBrearty, J.F., Marston, A.F. and Kanfer, F. H. Conditioning a verbal operant in a group setting: Direct vs. vicarious reinforcement. American Psychologist, 1961, 16, 425-434.

McDonald, F. J., Allen, D. W., and Orme, M. E. J. The effects of self-evaluation and social reinforcement on the acquisition of a teaching behavior. Paper delivered at American Educational Research Association Convention. Chicago: February 1966.

McNemar, Q. Psychological statistics. (3rd ed.) New York: Wiley, 1962.

Michael, D. N., and Maccoby, N. Factors influencing verbal learning from films under conditions of audience participation. In Lumsdaine, A. A., and Glaser, R. (Eds.), Teaching machines and programmed learning. Washington: National Education Association, 1960, Appendix One,665.

Miller, N E., and Dollard, J. Social learning and imitation. New Haven: Yale University Press, 1941.

Mowrer, O. H. Learning theory and the symbolic processes. New York: Wiley, 1960.

Porter, D. A. A critical review of a portion of the literature on teaching devices. Harvard Educational Review, 1957, 27, 126-147.

Rao, C. R. Linear statistical inference and its application. New York: Wiley, 1966.

Roshal, S. M. Film-mediated learning with varying representation of the task. In Lumsdaine, A. A. (Ed.), Student response in programmed instruction. Washington: National Research Council, 1961, 155-176.

Scheuler, H., Gold, M. J., and Mitzel, H. E. The use of television for improved teacher training and for improving measures of student-teaching performance: Phase I. Improvement of student teaching. Hunter College of the City University of New York. United States Office of Education Grant No. 730035, 1962.

Sheffield, F. D. Theoretical considerations in the learning of complex sequential tasks from demonstration and practice. In Lumsdaine, A. A. (Ed.), Student response in programmed instruction. Washington: National Research Council, 1961, 117-140.

Sheffield, F. D., and Maccoby, N. Summary and interpretation of research on organizational principles in constructing filmed demonstrations. In Lumsdaine, A. A. (Ed.), Student response in programmed instruction. Washington: National Research Council, 1961, 117-140.

Stern, G. C. Measuring noncognitive variables in research on teaching. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand, McNally, 1963, 398-447.

Thelen, H. A. Work-emotionality theory of the group as an organism. In Koch, S. (Ed.), Psychology: A study of a science: Vol. III. Formulations of the person and the social context. New York: McGraw-Hill, 1959, 544-611.

Tolman, E. C. Principles of purposive behavior. In Koch, S. (ed.), Psychology: A study of a science. New York: McGraw-Hill, 1959, Vol. 2, 92-157.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Withall, J. Development of a technique for the measurement of socio-emotional climate in classrooms. Journal of Experimental Education, 1949, 17.

Withall, J. An objective measurement of a teacher's classroom interactions. Journal of Educational Psychology, 1956, 47, 203-212.

Wulff, J. J., and Kraeling, D. Familiarization procedures used as adjuncts to assembly-task training with a demonstration film. In Lumsdaine, A. A., (Ed.), Student response in programmed instruction. Washington: National Research Council, 1961, 141-154.